Теория Вероятностей - 24, МатФак ВШЭ

Андрей В. Дымов и Андрей В. Хохлов

15 октября 2024 г.

1 Предварительно о теории вероятностей

Имеется значительное число учебников по Теории Вероятностей и никак нельзя сказать, что все они копируют друг друга, хотя набор основных тем обычно практически неизменен. Как правило в них в начале представлены сюжеты элементарной (конечной и дискретной) теории с (как бы взятыми из реальности) интерпретациями ее применений, далее обсуждают предельные случаи некоторых конечных моделей, теоретико-множественные конструкции— алгебры подмножеств и теорию меры, теорию интеграла Лебега на пространствах с мерой, разбор основных примеров сходимостей последовательностей случайных моделей[1]. При начальном изучении разница в подходах, может спровоцировать восприятие теории вероятностей как чисто-абстрактного, формального знания или, напротив, как набора эмпирически-эвристических рецептов, которые непонятно как выстраивать логически. При этом, в Теорию Вероятностей часто включают и чисто-практические фрагменты Математической Статистики.

Поэтому представляется разумным в самом начале сказать несколько слов, проясняющих хотя бы некоторые существующие расхождения во взглядах на предмет. В целом от теории вероятностей (по исторической традиции) ждут содержательных и понятных интерпретаций в разных областях человеческой деятельности с одной стороны, и в то же время строгого в рамках традиционных требований к математическим сюжетам изложения. При рассмотрении конечных и дискретных моделей сделать это было сравнительно просто, но оказалось невозможным одними только конечными вероятностными моделями представить, например, результаты измерений реальных физических процессов. Переход от конечных моделей к использующим актуальные бесконечности оказался весьма трудоемким и некоторый прогресс был достигнут лишь к 30-м годам XX века.

Неоднозначность вероятностной модели мира

Термин «вероятность» часто употребим при описании наблюдений за окружающим миром, но при этом вопрос можно ли *непосредственно* наблюдать вероятность обычно даже не формулируется. В лучшем случае предлагается приблизительная трактовка того, какие именно действия надо предпринять, чтобы *приблизительно(!)* сосчитать некую величину между 0 и 1. Но при этом в естественнонаучных теориях понятие вероятности (и в квантовой теории понятие амплитуды вероятности) используется уже в *точных* формулах — эта традиция несоответствия, похоже, не имеет признаков к изменению. По-видимому,

[1] В зависимости от вкусов автора учебника упор в изложении делается либо на интерпретациях (и таким образом курс приобретает как бы инженерно-прикладной характер – см. например, известный курс для слушателей Академии им. Жуковского Е.С.Вентцель "Теория Вероятностей"), либо на абстрактных, чисто математических утверждениях о пространствах с мерой (см. например, А.А.Боровков "Теория Вероятностей")

предполагается, что Теория Вероятностей как математическая абстрактная дисциплина отвечает за все.

Если мы посмотрим чуть внимательней на проблему измерений реальных физических процессов, то легко увидим обстоятельства, которые послужили причиной тому, что в настоящее время существуют *различные* теории вероятностей, причем количественные выводы в каждой из этих теорий иногда не вполне ясно как сравнивать.

Во-первых, возможные разночтения состоят в формализации самого понятия **вероятности**: относится ли оно к единичному измерению или имеет смысл лишь для потенциально бесконечной *последовательности* измерений. Для описания свойств последовательностей измерений необходимо придать строгий смысл другому важному общеупотребительному термину – **независимости** измерений.

Второе обстоятельство состоит в том, что в окружающем мире для физических теорий всегда приняты масштабные ограничения — так, например, почти все согласны, что физические описания микромира специфичны и эти отличия не удается описать как простой дополнительный набор настраиваемых параметров в единой общей теории. Возникает следующий вопрос о строении теории: а можно ли надеяться, что соответствующие математические определения вероятности и независимости должны быть одинаковыми во всех случаях приложений теории, не зависеть от масштаба описываемого реального эксперимента?

Пока что просто перечислим основные современные направления, выросшие из исторического общеинтуитивного подхода в области конечных моделей:

- аксиоматический подход А.Н.Колмогорова (который единственно и будет в центре нашего внимания в дальнейшем), в рамках которого вероятность понимается как неотрицательная вещественнозначная мера $0 \leqslant \mathbb{P} \leqslant 1$ на подходящем множестве, а независимость определяется как соотношение на возникающие числовые значения.

- фриквентистский (частотный) подход Рихарда фон Мизеса, в рамках которого вероятность относится к повторениям измерений и определяется как свойство числовых рядов, этот подход (хотя и неявно) используется в Математической статистике. Заметим, что некоторые теоремы колмогоровской теории являются аксиомами для фриквентистов. Расхождение фриквентистов с колмогоровской теорией в общих чертах касается вопроса, к чему должна относиться вещественнозначная мера $0 \leqslant \mathbb{P} \leqslant 1$: к повторениям эксперимента или к единичному эксперименту.

- эвристический и свободный от какой-либо теоретико-множественной схемы байесианский подход[2], который однако же в конкретных задачах математической статистики иногда быстро приводит к ответам.

- квантовая теория вероятностей, которая фактически (но неявно) изначально использовалась в приложениях к задачам микромира, но, как оказалось, эта теория не сводима к колмогоровской аксиоматике и по существу ей противоречит. Специфика данной теории в том, что в ней присутствует линейная структура на рассматриваемых множествах.

- экзотические неколмогоровские теории, в которых рассматриваемая мера $\mathbb{P}$ не обязательно вещественнозначная, а например, комплекснозначная или p-адическая.

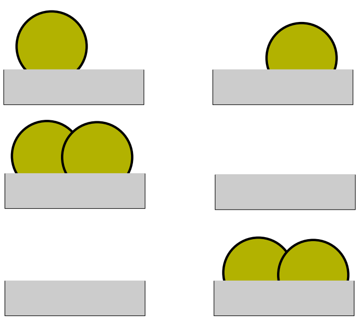
[2]В котором допустимо делать количественные вероятностные утверждения не только о данных в измерениях, подверженных случайным воздействиям. Например, фразы типа «Вероятность того, что Альберт Эйнштейн выпил чашку чая 1 августа, 1948 г. составляет 0.35» байесианцами считаются осмысленными, так как отражают частную силу веры в истинность предложения.

Пока что скажем, что колмогоровская теория вероятностей содержательно состоит из математических утверждений теории меры $\mathbb{P}$ применительно к аксиоматически введенному понятию независимости мер.

Необходимость теории

Само по себе понятие вероятности исторически казалось интуитивно ясным и использовалось для характеристики разных возможных ситуаций — это так называемая «оценка шансов». Однако желание иметь такую характеристику в численном выражении, с самого начала порождает трудности, известные под названием парадоксов теории вероятностей. Проиллюстрируем такую ситуацию следующим интуитивно понятным (но по вариантам ответов парадоксальным) вопросом: *В двух карманах случайно лежат две монеты, какова вероятность что монет в карманах поровну*

Подсчет всех шансов в эксперименте «монеты в карманах» указывает как минимум на два разных подхода: см. левую и правую картинку на Рис.1.



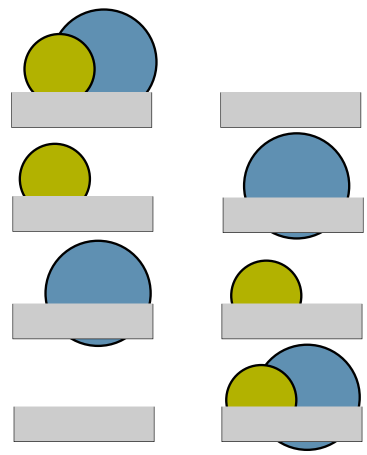


Рис. 1: Два способа подсчета шансов в эксперименте «монеты в карманах»

Интуитивный подход к оценке «вероятности того, что монет поровну» связан со сравнением количества шансов, составляющих «благоприятное» событие, с количеством всех «шансов» . Однако здесь кроется явная неоднозначность численного ответа: априори неясно, какую именно модель «всех шансов» следует избрать для приведенной словесной формулировки, да и роль симметрии разных шансов совершенно неясна. Важный вывод: *численные оценки вероятности имеют смысл только в связке с конкретно выбранной моделью эксперимента*, а без выбора модели само по себе утверждение о численном значении вероятности представляется лишенным смысла.

Поскольку задача количественного описания решается только с предъявлением модели вычисления вероятности (например, множества исходов, если мы хотим оставаться в рамках языка теории множеств), то для решения прикладных задач необходимо правило создания вероятностных моделей. Такого единого правила нет, и существующие разные версии теории вероятностей возникли отчасти поэтому. Зафиксированное множество «шансов» принято называть **пространством элементарных событий данного эксперимента**, в его выборе обычно имеется некоторый произвол (сравните с известным вопросом какой стороной выпадает монета при подбрасывании: шутники предлагают рассматривать случаи вставания монеты на ребро и зависания монеты в воздухе).

Небольшой модификацией вопроса про карманы и монеты можно пояснить базовую идею байесовского подхода: пусть карманы различаются как правый и левый и вопрос состоит в том, как количественный ответ зависит от априорного знания «правый карман непуст». Байесовский подход вообще избегает предъявления пространства элементарных событий, вместо этого изучаются зависимости ответа от описания эксперимента.

Интерпретации теории

Исторически так уж сложилось, что утверждения теории вероятностей широко используются как в самых разных приложениях так и в теоретической физике. Но надо признать, что в силу той же традиции описания моделей приводятся не всегда[3]. При практических рассмотрениях подбирают подходящую математическую модель, то есть пространство с мерой такое, что возможны количественные оценки вероятностей результатов экспериментов. Вопрос об оптимальном подборе такой модели не является частью теории, но зато *уже в рамках выбранной модели можно ставить вопрос о количественном соответствии реальности и вычисляемой в модели вероятности*: вообще говоря, процедура соответствия основана на повторениях экспериментов (измерений), методы проверки соответствия составляют отдельную дисциплину, которая называется Математической Статистикой.

При изложении любой теории возникает список основных понятий. План этого вводного курса теории вероятностей состоит в том, чтобы проследить за формулировками, начав с достаточно простых интуитивно ясных моделей и дополняя соответствующие построения до более общих случаев. Можно надеяться, что на этом пути связь аксиоматически вводимых понятий с интуитивным их пониманием (возникшем еще до всяческих теорий в процессе практической деятельности) станет более понятна.

2 Колмогоровская аксиоматика теории вероятностей

С этого момента и до конца нашего курса мы будем заниматься колмогоровской аксиоматической теорией вероятностей. Однако следует помнить, что помимо колмогоровского принципа используются и другие (неколмогоровские) теории, причем их использование весьма распространено — мы коротко поговорим о них чуть позже.

Пусть имеется эксперимент, исходы которого мы хотели бы интерпретировать как случайные. Наша цель — вычислить какие-нибудь вероятности, связанные с этими исходами. Например, мы бросаем два игральных кубика, на гранях каждого из которых написаны числа от 1 до 6, и хотим найти вероятность того, что сумма результатов двух бросков делится на 3. При этом заметим, что понятие "вероятность" пока не определено.

Возникает две задачи. Первая — формулировка математической модели, описывающей этот эксперимент и, в частности, определение понятия вероятность. Вторая — непосредственное вычисление интересующих нас вероятностей в рамках построенной модели.

В колмогоровской аксиоматике математическая модель состоит из трех компонент: множество исходов Ω нашего эксперимента (выше мы его называли множество "шансов"), сигма-алгебра его подмножеств $\mathcal{F}$ (вероятность только этих подмножеств и окажется возможным измерить), и вероятность $\mathbb{P}$, являющаяся мерой на измеримом пространстве $(\Omega, \mathcal{F})$. При этом в случае, когда множество Ω не более, чем счетно, можно выбрать $\mathcal{F} = 2^{\Omega}$ и исключить сигма-алгебру $\mathcal{F}$ из рассмотрения. Далее мы обсуждаем обозначенные компоненты детально.

[3]Например, современные попытки аксиоматически сформулировать принципы квантовой механики или квантовой теории поля включают аксиомы, в которые вероятность входит как готовый параметр теории, что это означает с точки зрения постановки эксперимента часто оставляют без разъяснений.

2.1 Множество элементарных исходов Ω

Колмогоровский подход предполагает, что любой эксперимент можно описать в терминах набора альтернатив ω, называемых *элементарными исходами*. Их объединение составляет некоторое множество Ω, называемое *множеством элементарных исходов*. Для каждого эксперимента множество элементарных исходов Ω выбирается индивидуально, причем этот выбор зависит от того, что именно вас интересует в рассматриваемом эксперименте, и основан на вашем личном жизненном опыте и вашей интуиции. Математика же начинается *только* с того момента, когда множество Ω зафиксировано.

Несмотря на то, что Ω также называют *пространством элементарных исходов*, на Ω *не* предполагается наличия никакой структуры, в частности не требуется структуры линейного пространства.

Пример 2.1. Эксперимент состоит в подбрасывании одной монеты $n \geq 1$ раз. Нас интересует, что выпадает при каждом броске, орел или решка. Разумно выбрать

$$\Omega = \{"\text{О}", "\text{Р}"\}^n = \big\{(i_1, \dots, i_n) : \, i_j = "\text{О}" \text{ или } "\text{Р}", \, j = 1, \dots, n\big\}. \tag{2.1}$$

Таким образом, элементарными исходами нашего эксперимента являются всевозможные вектора $(i_1, \dots, i_n) \in \Omega$, где $i_j =$"О" , если на j-ом броске выпал орел, и $i_j =$"Р" в противном случае.

Но, если вы ни разу в жизни не видели, чтобы при подбрасывании монеты выпадала решка, и вы не верите людям, которые говорят, что решка иногда таки выпадает, вам было бы разумно выбрать

$$\Omega = \{0\}. \tag{2.2}$$

Здесь $\omega = 0$ отвечает единственному рассматриваемому вами исходу эксперимента, когда выпало n орлов подряд. [4]

Пример 2.2. Множество элементарных исходов (2.1) также описывает эксперимент, состоящий в одновременном подбрасывании n *пронумерованных* монет. Так как монеты пронумерованы, они *различимы*. В частности, при $n = 2$ построенное множество элементарных исходов можно записать в виде

$$\Omega = \{"\text{ОО}", "\text{ОР}", "\text{РО}", "\text{РР}"\}. \tag{2.3}$$

Такое множество Ω схематически изображено на правой части рис. 1.

В ситуации же, когда монеты *неразличимы*, изображенной на левой части рис. 1, разумно выбрать

$$\Omega = \{"\text{ОО}", "\text{ОР}", "\text{РР}"\}.$$

При такой модели эксперимента мы можем говорить только о количестве выпавших орлов и решек, но не о последовательности в которой они выпадали, так что элементарные исходы "ОР" и "РО" из (2.3) в этом случае отождествляются.

Пример 2.3. Рассмотрим упоминавшийся выше эксперимент с бросанием двух кубиков. Если считать кубики различимыми, то разумно выбрать

$$\Omega = \big\{(i, j) : \, 1 \leq i, j \leq 6\big\}, \tag{2.4}$$

[4] Несмотря на все "реверансы" , которые мы делаем и будем делать по поводу субъективности выбора математической модели рассматриваемого эксперимента, на контрольных мероприятиях выбор (2.2) мы вряд ли оценим, хотя с *математической* точки зрения он вполне корректен.

где i — число, выпавшее на первом кубике, а j — на втором. Если считать их неразличимыми, то

$$\Omega = \big\{(i, j) : 1 \le i \le j \le 6\big\}, \tag{2.5}$$

Если же теперь вспомнить, что нас интересовала лишь сумма выпавших очков, при этом нам не важно как именно сумма разбита по двум кубикам, мы можем выбрать

$$\Omega = \big\{2, 3, \ldots, 11, 12\big\}. \tag{2.6}$$

Пример 2.4. Вы знаете, что к вам зайдет знакомый между 8 и 9 часами вечера, и никакой дополнительной информации у вас нет. Разумно считать время его визита случайным числом между 8 и 9 часами. Выбора $\Omega = [0, 60] \subset \mathbb{R}$ хватит для любой точности измерений. Однако, если мы округляем время до минуты, то можно положить $\Omega = \{0, 1, \ldots, 60\}$. А можно и оставить прежний выбор $\Omega = [0, 60]$, просто нецелые элементарные исходы $\omega \in \Omega$ никогда не реализуются. Заметим, что обычно в подобных задачах рассматриваемый отрезок сжимают до отрезка единичной длины, выбирая $\Omega = [0, 1]$.

2.2 Вероятностное пространство в случае не более, чем счетного множества Ω

В примере 2.4 предлагается рассматривать континуальное множество элементарных исходов Ω. Этот случай заметно сложнее, чем ситуация, рассмотренная во всех остальных примерах, в которых множество Ω было не более, чем счетным. Вскоре мы вернемся к общему случаю, а пока, для упрощения восприятия, сделаем

Предположение 2.5. Множество элементарных исходов Ω конечно либо счетно.

Рассмотрим функцию $p : \Omega \mapsto [0, 1]$, удовлетворяющую свойству

$$\sum_{\omega \in \Omega} p(\omega) = 1. \tag{2.7}$$

Число $p(\omega)$ будем называть *вероятностью элементарного исхода* ω.

Событием или *случайный событием* мы будем называть произвольное подмножество $A \subset \Omega$. Эта терминология довольно естественна: событие состоит из набора элементарных исходов. Так, в примере 2.2 можно рассмотреть событие $A = \{$"ОР" , "РО" $\}$, состоящее в том, что выпали один орел и одна решка.

Наконец, определим функцию множеств $\mathbb{P} : 2^{\Omega} \mapsto [0, 1]$, сопоставляющую каждому событию $A \subset \Omega$ число по правилу [5]

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega). \tag{2.8}$$

Функция $\mathbb{P}$ называется *распределением вероятностей*, или *вероятностной мерой*, или просто *вероятностью*, а число $\mathbb{P}(A)$ — *вероятностью события* A. Отметим, что $\mathbb{P}$ действительно является счетно-аддитивной мерой.

Из определения сразу вытекает следующий результат (докажите его!):

Лемма 2.6. *1)* $\mathbb{P}(\emptyset) = 0$*;*

2) $\mathbb{P}(\Omega) = 1$*;*

3) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *для любого события* $A \subset \Omega$ *и его дополнения* $A^c := \Omega \setminus A$*;*

[5] Так как Ω не более, чем счетно и $p(\omega) \ge 0$, сумма ряда (2.8) хорошо определена, то есть конечна и не зависит от порядка суммирования.

3) Для любых событий $A, B \subset \Omega$,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

В частности, если $A \cap B = \emptyset$,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Традиционно выделяют ситуацию, особенно часто встречающуюся в приложениях, когда вероятности $p(\omega)$ всех элементарных исходов $\omega \in \Omega$ равны. Тогда, согласно требованию (2.7), находим

$$p(\omega) = \frac{1}{|\Omega|} \quad \forall \omega \in \Omega \quad \text{и} \quad \mathbb{P}(A) = \frac{|A|}{|\Omega|} \quad \forall A \subset \Omega.$$

Такое определение вероятности называют *классическим*.

Наконец, мы готовы дать основное определение, которое однако работает *только в рамках предположения 2.5*.

Определение 2.7 (случай не более, чем счетного Ω)**.** Пара $(\Omega, \mathbb{P})$ называется *вероятностным пространством*.

! Выбрать вероятностное пространство и означает построить математическую модель вашего эксперимента **!**

Важно очень хорошо понимать, что выбор вероятностного пространства субъективен, связан с вашей "физической" интуицией, и не может быть правильным или неправильным, потому что критерий правильности не определен. Интересуясь вопросом "с какой вероятностью во время утренней пробежки я встречу динозавра" , вполне можно положить $\Omega = \{$"Да" , "Нет" $\}$ с вероятностями элементарных исходов $p($"Да" $) = p($"Нет" $) = 1/2$. Нельзя утверждать, что этот выбор неправильный, но можно лишь сказать, что он плохо соответствует жизненному опыту большинства из нас. Выбор вероятностного пространства — это задание аксиоматики, в рамках которой вы планируете работать.

Легко провести аналогию с геометрией: пока вы не задали аксиоматику, нельзя ответить на вопрос "верно ли, что через точку, не лежащую на данной прямой, можно провести лишь одну прямую, параллельную данной?" . Действительно, в евклидовой геометрии, к которой большинство из нас приучил жизненный опыт, ответ на этот вопрос положителен, а в геометрии Лобачевского — отрицателен. Одну из этих геометрий *разумно* применять в одних задачах, другую — в других. *Математика* же начинается *только* после того, как задана аксиоматика. В колмогоровской теории вероятностей — как только введено вероятностное пространство $(\Omega, \mathbb{P})$. И только после этого утверждения "правильно" или "неправильно" обретают смысл.

Итак, решение любой "текстовой" задачи по теории вероятностей состоит из двух этапов. Первый этап не имеет прямого отношения к математике и состоит в выборе вероятностного пространства; это может быть сделано разными способами. Второй этап состоит в решении *математической* задачи в рамках выбранной модели.

Пример 2.8. В примере 2.1 имеем $|\Omega| = 2^n$. Следовательно, классическое определение вероятности дает

$$p(\omega) = \frac{1}{2^n} \quad \forall \omega \in \Omega \quad \text{и} \quad \mathbb{P}(A) = \frac{|A|}{2^n} \quad \forall A \subset \Omega.$$

Такой выбор разумен, если вы считаете, что ваша монетка симметрична. А теперь обдумайте как определить вероятностное пространство в случае, когда монетка, которую мы бросаем, такова, что вероятности выпадения орла и решки при единичном броске не равны. Пока вы размышляете, мы двинемся дальше, но обязательно вернемся к этому сюжету позже, когда введем ключевое понятие *независимости событий*. Отметим только некорректность предложенного вопроса: мы начали говорить о вероятности выпадения орла либо решки *до* определения вероятностного пространства.

Пример 2.9. Вернемся к задаче о подбрасывании двух кубиков, которую мы начали обсуждать в начале параграфа 2 и продолжили в примере 2.3. Напомним, что нас интересует вероятность события

$$A = \{\text{"сумма результатов двух бросков кратна 3" }\}. \tag{2.9}$$

а) В случае различимых кубиков, когда множество элементарных исходов выбрано в виде (2.4), получаем, что событие A состоит из элементарных исходов $(1,2),(1,5),(2,4),(3,6)$, симметричных им исходов $(2,1),(5,1),(4,2),(6,3)$, а также исходов $(3,3),(6,6)$. Таким образом, $|A| = 10$, а $|\Omega| = 36$, так что классическое определение вероятности дает

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{5}{18}. \tag{2.10}$$

б) В случае, когда кубики считаются неразличимыми, то есть множество элементарных исходов имеет вид (2.5), получаем $A = \{(1,2),(1,5),(2,4),(3,6),(3,3),(6,6)\}$, так что $|A| = 6$. Так как $|\Omega| = 21$, классическое определение вероятности дает

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{2}{7}. \tag{2.11}$$

в) В ситуации, когда множество элементарных исходов выбрано в виде (2.6), $|\Omega| = 11$, $A = \{3,6,9,12\}$, так что классическое определение вероятности дает

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{4}{11}. \tag{2.12}$$

Как видно из последнего примера, разные выборы моделей ведут к разным вероятностям одного и того же события. Возможно, в этом месте читатель захочет закрыть конспект со словами "нет, ну это чушь, конечно единственный правильный выбор вероятностного пространства сделан в пункте (а), и вообще не понятно, где может возникнуть ситуация неразличимых объектов" . Действительно, если у нас есть два совершенно одинаковых, неразличимых кубика, можно один покрасить в синий цвет, а второй — в красный, и конечно вероятность события (2.9) от этого не изменится. Ну, а модель (в) не имеет вообще никакого отношения к реальности, так как, конечно, сумма будет равна 6 гораздо чаще, чем 12. Однако, все эти разговоры имеют отношение лишь к соответствию выбранной модели окружающей нас действительности и нашей интуиции в отношении слова "вероятность" , в то время как с точки зрения математики все три модели *совершенно* корректны. Более того, к интуиции нужно относиться очень аккуратно. Вспоминая известный феномен кота Шрёдингера, мы убеждаемся, что даже наблюдение за системой, уж не говоря о раскраске ее в разные цвета, далеко не невинное действие. А рассуждая о соответствии действительности модели пункта (в), мы отождествили частоту встречаемости события при повторном эксперименте с его вероятностью при единичном, но это совершенно разные объекты. Результат же, утверждающий, что при увеличении числа независимых повторений эксперимента частота, с которой происходит данное событие, стремится к его вероятности при единичном эксперименте, действительно имеет место и называется *закон больших чисел (ЗБЧ)*. Это один из центральных результатов нашего курса, он требует уточнения и доказательства (в отличие от фриквентистского подхода, в котором ЗБЧ постулируется и фактически является определением вероятности).

Завершим это обсуждение следующим примером, подтверждающим наш скепсис относительно интуиции, особенно когда дело касается непривычных нам условий; в данном случае — непривычных нам масштабов микромира.

Пример 2.10. Распределение n бозонов по m уровням энергии [6] управляется так называемой статистикой Бозе-Эйнштейна. Она соответствует классическому определению вероятности в задаче о размещении n *неразличимых* частиц по m *различимым* ячейкам. В этом случае множество элементарных исходов можно выбрать в виде

$$\Omega = \{\omega = (n_1, \dots, n_m) : n_1 + \dots + n_m = n\}.$$

Событие $A_{i,k}$, состоящее в том, что в ячейке номер i содержится ровно k частиц, имеет вид $A_{i,k} = \{\omega : n_i = k\}$. С помощью стандартного метода "шаров и перегородок" (см. аппендикс A.2.1) можно проверить, что $|A_{i,k}| = C^n_{n+m-1}$ и $|\Omega| = C^{n-k}_{n+m-k-2}$, так что

$$\mathbb{P}(A) = \frac{C^n_{n+m-1}}{C^{n-k}_{n+m-k-2}}.$$

Резюмируем. Пусть вам поставили следующую задачу: имеется 6 ящиков, в которые случайным образом разбросали 15 шаров. Требуется найти вероятность, что останутся пустые ящики. Прежде, чем ее решать, необходимо выяснить, предполагаются ли шары различимыми, и предполагаются ли ящики различимыми. Очевидно, имеется 4 варианта, каждый из которых требует своего решения задачи, см. Аппендикс A.2.1.

2.3 Вероятностное пространство: общий случай

В случае, когда множество элементарных исходов Ω несчетно (как в примере 2.4), определение вероятности, данное в (2.7), (2.8), очевидно не работает. Поэтому предлагается его нижеследующее обобщение, основанное на теории меры. См. Аппендикс B, в котором мы напоминаем некоторые базовые факты теории меры.

Предположим, что на множестве Ω задана сигма-алгебра его подмножеств $\mathcal{F}$.

Определение 2.11. Неотрицательная мера $\mathbb{P}$ на измеримом пространстве $(\Omega, \mathcal{F})$ называется *вероятностной мерой* или *вероятностным распределением*, если $\mathbb{P}(\Omega) = 1$.

Следующий результат немедленно следует из стандартных свойств меры и тождества $\mathbb{P}(\Omega) = 1$ (проверьте!).

Лемма 2.12. *Верны утверждения леммы 2.6, где $A, B \in \mathcal{F}$.*

Теперь дадим основное определение вероятностного пространства, работающее для произвольного множества элементарных исходов Ω (в отличие от определения 2.7). Если вас просят дать определение вероятностного пространства, нужно приводить именно его.

Определение 2.13. *Вероятностным пространством* называется тройка $(\Omega, \mathcal{F}, \mathbb{P})$, где Ω — множество элементарных исходов, $\mathcal{F}$ — σ-алгебра его подмножеств, а $\mathbb{P}$ — вероятностная мера на измеримом пространстве $(\Omega, \mathcal{F})$.

Определение 2.7 является частным случаем определения 2.13. Действительно, легко проверить, что функция множеств (2.8) задает вероятностную меру на измеримом пространстве $(\Omega, \mathcal{F})$, где $\mathcal{F} = 2^\Omega$ — σ-алгебра всех подмножеств (не более, чем счетного) множества Ω.

[6] При микроканоническом ансамбле.

Пример 2.14. Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}[0,1], Leb)$, где $\mathcal{B}[0,1]$ обозначает σ-алгебру борелевских подмножеств отрезка $[0,1]$, а Leb обозначает меру Лебега на $([0,1], \mathcal{B}[0,1])$. Такое вероятностное распределение $\mathbb{P}$ называется *равномерным распределением* на отрезке $[0,1]$. Предложенная вероятностное пространство разумно для примера (2.4).

Отметим, что в примере 2.14 $\mathbb{P}\{\omega\} = 0$ для каждого элементарного исхода $\omega \in \Omega$. Это характерно для многих вероятностных моделей с несчетным множеством элементарных исходов (в отличие от случая счетного множества Ω, см. (2.7)).

Подчеркнем появление σ-алгебры $\mathcal{F}$ в определении 2.7, в отличие от определения 2.13. Оно обусловлено тем, что область определения меры $\mathbb{P}$ не обязана содержать все подмножества множества Ω. Уже в примере 2.14 это не так: известно, что для меры Лебега существуют неизмеримые подмножества, поэтому приходится ограничить область определения $\mathbb{P}$ сигма-алгеброй $\mathcal{F} = \mathcal{B}[0,1]$. [7] Таким образом, интерпретация σ-алгебры $\mathcal{F}$ такова: она задает набор множеств, вероятность которых мы можем измерить.

Отметим естественность требования, что область определения $\mathcal{F}$ вероятности является алгеброй. К примеру, если имеются два события $A, B \in \mathcal{F}$, естественно задать вопросы "с какой вероятностью произойдет хотя бы одно из событий A и B" и "с какой вероятностью произойдут оба события A и B" . Другими словами, требуется найти вероятности $\mathbb{P}(A \cup B)$ и $\mathbb{P}(A \cap B)$.

Свойство $\mathcal{F}$ быть не только алгеброй, но и σ-алгеброй, а вероятности $\mathbb{P}$ быть σ-аддитивной, также оказывается весьма полезным. В частности, лемма B.9(4) утверждает следующий естественный факт (неверный без свойства σ-аддитивности меры $\mathbb{P}$!): если имеется последовательность "уменьшающихся" событий $C_i \in \mathcal{F}$, то есть таких, что $C_{i+1} \subset C_i$, и их вероятности $\mathbb{P}(C_i)$ стремятся к нулю при $i \to \infty$, то вероятность того, что они все произойдут сразу равна нулю:

$$\mathbb{P}(\bigcap_i C_i) = 0.$$

Это свойство называется *непрерывностью вероятности и нуле.*

Задача 2.15. 1. Проинтерпретируйте пункты (2,3) леммы B.9 с точки зрения теории вероятностей, аналогично тому, как это сделано выше для пункта (4).
2. Пусть μ — неотрицательная аддитивная функция множеств на измеримом пространстве $(\Omega, \mathcal{F})$. Докажите, что для любых не пересекающихся друг с другом множеств $C_i \in \mathcal{F}$, $i \geq 1$, выполнено

$$\mu(\bigcup_{i=1}^{\infty} C_i) \geq \sum_{i=1}^{\infty} \mu(C_i).$$

3 Условная вероятность и независимость

3.1 Условная вероятность

Рассмотрим следующую задачу.

Пример 3.1. В семье имеется два (различимых!) ребенка. Какова вероятность того, что оба — девочки? Разумным выбором вероятностного пространства для этой задачи

[7] В этом примере мы могли бы выбрать вместо борелевской сигма-алгебры более широкую σ-алгебру подмножеств, измеримых по Лебегу. Но для целей теории вероятностей обычно хватает выбора $\mathcal{F} = \mathcal{B}[0,1]$.

представляется $(\Omega, \mathcal{F}, \mathbb{P})$, где

$$\Omega = \{\text{ММ, МД, ДМ, ДД}\}, \qquad \mathbb{P}\{\omega\} = \frac{1}{|\Omega|} = \frac{1}{4} \quad \forall \omega \in \Omega, \qquad \mathcal{F} = 2^{\Omega}. \tag{3.1}$$

Тогда интересующее нас событие имеет вид $A = \{\text{ДД}\}$, так что $\mathbb{P}(A) = 1/4$.

Зададимся теперь вопросом, какова вероятность того, что оба ребенка — девочки, при условии, что хотя бы один ребенок — девочка? [8] Введенная в предыдущем параграфе аксиоматика не дает никакого рецепта как можно было бы учесть это условие в рамках предложенной в примере 3.1 модели. Разумным представляется построить новое вероятностное пространство $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, учитывающее это условие:

$$\widetilde{\Omega} = \{\text{МД, ДМ, ДД}\}, \qquad \widetilde{\mathbb{P}}\{\omega\} = \frac{1}{|\Omega|} = \frac{1}{3} \quad \forall \omega \in \Omega, \qquad \mathcal{F} = 2^{\widetilde{\Omega}}. \tag{3.2}$$

В этом случае интересующее нас событие $A = \{\text{ДД}\}$ имеет вероятность $\mathbb{P}(A) = 1/3$.

Однако, такой подход неудобен. Если мы захотим поменять рассматриваемое условие на другое (например, "второй ребенок — мальчик"), нам придется строить новое вероятностное пространство. В приложениях же (и математических тоже) регулярно требуется вычислять вероятность одного и того же события при множестве различных предположений. Например, страховую компанию может интересовать вероятность события A_I, что автомобиль попадет в ДТП с ущербом, лежащем в интервале $I \subset \mathbb{R}$ рублей. При этом разумно оценивать не только (безусловную) вероятность $\mathbb{P}(A_I)$, но и вероятности события A при условии, что автомобиль был той или иной марки, отдельно для каждой марки автомобиля. Следующее определение позволяет обойтись одним и тем же вероятностным пространством при вычислении всевозможных вероятностей со всевозможными условиями.

Рассмотрим произвольное вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и события $A, H \in \mathcal{F}$, причем $\mathbb{P}(H) > 0$.

Определение 3.2. *Условной вероятностью события A при условии события $H, \mathbb{P}(H) >$* 0, называется число [9]

$$\mathbb{P}(A|H) = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)}. \tag{3.3}$$

Если $\mathbb{P}(H) = 0$, то условная вероятность $\mathbb{P}(A|H)$ не определена. Общественный договор таков: если вас просят вычислить вероятность какого-либо события A при зафиксированных дополнительных данных H, всегда подразумевается условная вероятность в смысле определения 3.2. При этом формулировки могут быть совершенно различны. Вместо того, чтобы говорить об «условной вероятности того, что случайно выбранное лицо страдает дальтонизмом при условии, что это лицо — женщина», можно говорить о вероятности того, что женщина страдает дальтонизмом. Часто слова «при условии H» заменяют словами, «если известно, что H произошло». Короче говоря, наши формулы и символы не допускают никакой двусмысленности, но словесные выражения часто недостаточно четки и требуют точного истолкования. В противоположность условной вероятности $\mathbb{P}(A|H)$ говорят для ясности о безусловной вероятности $\mathbb{P}(A)$. Строго говоря, эпитет «безусловная» является лишним, и его можно опускать.

Убедимся, что определенная таким образом условная вероятность отвечает нашей интуиции.

[8] Другими словами: если известно, что хотя бы один ребенок — девочка.

[9] Нередко событие H называют *гипотезой.*

Пример 3.3. Пусть в примере 3.1 требуется вычислить вероятность того, что оба ребенка — девочки, при условии, что хотя бы один — девочка. Другими словами, требуется найти условную вероятность $\mathbb{P}(A|H)$, где $H = \{\text{МД, ДМ, ДД}\}$. По определению 3.2,

$$\mathbb{P}(A|H) = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}\{\text{ДД}\}}{\mathbb{P}(H)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Мы видим, что ответ совпадает с результатом, полученным ранее в рамках модели (3.2).

Пример 3.4. Рассмотрим дискретное вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, где Ω — не более, чем счетное множество, вероятность $\mathbb{P}$ задана формулами (2.7)-(2.8), а $\mathcal{F} = 2^\Omega$. Пусть $H \subset \Omega$. Тогда, согласно формуле (3.3), $\forall A \in \Omega$,

$$\mathbb{P}(A|H) = \sum_{\omega \in A} p_H(\omega), \quad \text{где} \quad p_H(\omega) = \begin{cases} \frac{p(\omega)}{\mathbb{P}(H)}, & \text{если} \quad \omega \in H, \\ 0, & \text{если} \quad \omega \notin H. \end{cases}$$

Заметим, что $\mathbb{P}(\Omega|H) = \sum_{\omega \in \Omega} p_H(\omega) = 1$.

Пример (3.4) проясняет смысл происходящего. В случае дискретного вероятностного пространства условная вероятность устроена точно так же, как и безусловная, то есть задается формулами (2.7)-(2.8), но вероятности элементарных исходов $p(\omega)$ модифицированы. А именно, вероятность исходов ω, противоречащих событию H (то есть, $\omega \notin H$), обнуляются, а вероятности остальных ω нормируются так, чтобы их сумма была равна 1.

В общем случае (3.3) ситуация аналогична: вероятность куска $A \setminus H$ события A, противоречащего H, обнуляется, а вероятность оставшегося куска $A \cap H$ нормируется мультипликативной константой так, чтобы $\mathbb{P}(\Omega|H) = 1$.

Теперь обсудим свойства условных вероятностей. Основная их часть следует из следующей тривиальной леммы.

Лемма 3.5. *Условная вероятность* [10] $\mathbb{P}(\cdot|H)$ *является вероятностной мерой на измеримом пространстве* $(\Omega, \mathcal{F})$.

Доказательство. Из определения 3.2 следует

- Неотрицательность: $\mathbb{P}(A|H) \geq 0 \ \forall A \in \mathcal{F}$,
- Нормировка: $\mathbb{P}(\Omega|H) = \dfrac{\mathbb{P}(H)}{\mathbb{P}(H)} = 1$,
- σ-аддитивность:

$$\mathbb{P}(\cup_{i=1}^\infty A_i|H) = \frac{\mathbb{P}\big(\cup_{i=1}^\infty (A_i \cap H)\big)}{\mathbb{P}(H)} = \frac{\sum_{i=1}^\infty \mathbb{P}(A_i \cap H)}{\mathbb{P}(H)} = \sum_{i=1}^\infty \mathbb{P}(A_i|H),$$

  для любого набора событий $A_i \in \mathcal{F}$, таких что $A_i \cap A_j = \emptyset$ при $i \neq j$.

□

В частности, каждое событие $H \in \mathcal{F}$ естественным образом порождает новое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P}(\cdot|H))$, отличающееся от старого выбором вероятности.

[10] Значок $\cdot$ в формуле $\mathbb{P}(\cdot|H)$ указывает на место, куда нужно подставить аргумент при вычислении значений рассматриваемой функции. То есть, $\mathbb{P}(\cdot|H)$ — функция, сопоставляющая множествам $A \in \mathcal{F}$ числа $\mathbb{P}(A|H)$.

Согласно лемме 3.5, условная вероятность $\mathbb{P}(\cdot|H)$ удовлетворяет всем утверждениям, касающимся безусловной вероятности $\mathbb{P}(\cdot)$. Например, для любых событий $A, B \in \mathcal{F}$,

$$\mathbb{P}(A \cup B|H) = \mathbb{P}(A|H) + \mathbb{P}(B|H) - \mathbb{P}(A \cap B|H).$$

Еще раз подчеркнем, что это равенство не требует отдельного доказательства, а следует из лемм 3.5 и 2.12. Отметим, однако, что сказанное выше касается лишь функции $\mathbb{P}(\cdot|H)$ при фиксированном множестве H.

Задача 3.6. Постройте пример вероятностного пространства $(\Omega, \mathcal{F}, \mathbb{P})$ и событий $A, H \in \mathcal{F}$, $\mathbb{P}(H) \neq 0$, таких что
а) $\mathbb{P}(A|H) + \mathbb{P}(A|H^c) \neq 1$,
б) $\mathbb{P}(A|H) + \mathbb{P}(A^c|H^c) \neq 1$.

Все конструкции, построенные для безусловной вероятности, можно применить и к условной. В частности, конструкцию условной вероятности:

Задача 3.7. Пусть $H_1, H_2 \in \mathcal{F}$. Обозначим $\mathbb{P}_{H_i} := \mathbb{P}(\cdot|H_i)$, $i = 1, 2$. Докажите, что

$$\mathbb{P}_{H_1}(\cdot|H_2) = \mathbb{P}(\cdot|H_1 \cap H_2) = \mathbb{P}_{H_2}(\cdot|H_1).$$

То есть, для любого $A \in \mathcal{F}$,

$$\mathbb{P}_{H_1}(A|H_2) = \mathbb{P}(A|H_1 \cap H_2) = \mathbb{P}_{H_2}(A|H_1).$$

3.2 Несколько очень полезных соотношений

В этом параграфе для пересечения событий $A_1 \dots, A_n \in \mathcal{F}$ мы будем использовать обозначение

$$A_1 \dots A_n := A_1 \cap \dots \cap A_n.$$

Все результаты этого параграфа совершенно тривиальны с математической точки зрения, но удивительно эффективны в приложениях.

Формула умножения вероятностей

Пусть $A, B \in \mathcal{F}$, причем $\mathbb{P}(A) \neq 0$. Тогда из определения условной вероятности (3.3) сразу следует соотношение

$$\mathbb{P}(AB) = \mathbb{P}(B|A)\mathbb{P}(A). \tag{3.4}$$

По индукции получаем *формулу умножения вероятностей*:

$$\mathbb{P}(A_1 \dots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \cdots \mathbb{P}(A_n|A_1 \dots A_{n-1}).$$

Формула полной вероятности

Будем говорить, что события $H_1, \dots, H_n$ задают *разбиение* события A, если $H_i \cap H_j = \emptyset$ при $i \neq j$ и $A \subset \cup_{i=1}^n H_i$.

Следующая формула является основным средством вычисления вероятности событий, естественным образом разбивающихся на несколько кусков.

Предложение 3.8. *Пусть события $H_1, \dots, H_n$ задают разбиение события A и имеют положительные вероятности $\mathbb{P}(H_i) > 0$. Тогда*

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(AH_i) = \sum_{i=1}^n \mathbb{P}(A|H_i)\mathbb{P}(H_i).$$

Доказательство. Первое равенство следует из аддитивности меры и соотношения $A = (AH_1) \cup \dots \cup (AH_n)$, $(AH_i) \cap (AH_j) = \emptyset$ для $i \neq j$. Второе равенство следует из первого и формулы (3.4). □

На практике события H_i обычно выбираются так, чтобы $\cup_{i=1}^n H_i = \Omega$. В этом случае говорят, что семейство H_i образует *полную группу событий*.

Пример 3.9. Имеется две урны, в одной 1 черный шар и 2 белых, а во второй по одному черному и белому шару. Вы наугад выбираете урну и вытаскиваете оттуда шар. Какова вероятность вытащить белый?

Сперва нужно построить вероятностное пространство, отвечающее этому эксперименту. Шары будем считать пронумерованными, так что множество элементарных исходов разумно выбрать в виде

$$\Omega = \{U_1B, U_1W_1, U_1W_2, U_2B, U_2W\},$$

где префикс "U_j" отвечает выбору j-ой урны, а B и W обозначают цвет выбранного шара: "black" или "white" .

Теперь нужно определить вероятности элементарных исходов. Условие задачи разумно интерпретировать следующим образом:

$$\mathbb{P}(U_1) = \mathbb{P}(U_2) = 1/2, \tag{3.5}$$

где $U_1 = \{U_1B, U_1W_1, U_1W_2\}$ — событие, что мы выбрали первую урну, а U_2 определено аналогично. Кроме того, "внутри" каждой урны вероятности распределены классическим образом:

$$\mathbb{P}(U_1B|U_1) = \mathbb{P}(U_1W_1|U_1) = \mathbb{P}(U_1W_2|U_1) = 1/3, \quad \mathbb{P}(U_2B|U_2) = \mathbb{P}(U_2W|U_2) = 1/2. \tag{3.6}$$

Согласно формуле умножения вероятностей (3.4), эти условия однозначно определяют вероятности элементарных исходов:

$$\mathbb{P}(U_1B) = \mathbb{P}(U_1W_1) = \mathbb{P}(U_1W_2) = 1/6, \quad \mathbb{P}(U_2B) = \mathbb{P}(U_2W) = 1/4.$$

Теперь конструкция, предложенная в параграфе 2.2, дает желаемое вероятностное пространство, и мы можем приступить к вычислению требуемой вероятности события

$$White := \{U_1W_1, U_1W_2, U_2W\}.$$

Это можно сделать просто сложив вероятности соответствующих элементарных исходов:

$$\mathbb{P}(White) = 1/6 + 1/6 + 1/4 = 7/12.$$

Однако, на практике все хорошо понимают, [11] что условия (3.5) и (3.6) однозначно задают вероятностное распределение, поэтому вопрос "существования" не возникает, и вероятностное пространство обычно явно не строят. Вместо этого напрямую вычисляют вероятность требуемого события, используя формулу полной вероятности:

$$\mathbb{P}(White) = \mathbb{P}(White|U_1)\mathbb{P}(U_1) + \mathbb{P}(White|U_2)\mathbb{P}(U_2) = 2/3 \cdot 1/2 + 1/2 \cdot 1/2 = 7/12.$$

[11] "Блажен, кто верует"

Формула Байеса

Пусть события $A, B \in \mathcal{F}$ имеют положительные вероятности. Тогда из формулы умножения вероятностей тривиальным образом следует *формула Байеса*:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{3.7}$$

Она имеет следующее развитие.

Предложение 3.10 (Теорема Байеса)**.** *Если события $H_1, \ldots, H_n$ задают разбиение множества A и $\mathbb{P}(H_j) > 0 \,\forall j$, $\mathbb{P}(A) > 0$, то*

$$\mathbb{P}(H_k|A) = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\sum_{j=1}^{n} \mathbb{P}(A|H_j)\mathbb{P}(H_j)}, \quad \forall k.$$

Доказательство: Первое равенство есть в точности формула Байеса (3.7). Чтобы получить второе равенство, мы применили формулу полной вероятности к знаменателю $\mathbb{P}(A)$. □

Несмотря на возможное ощущение пустого жонглирования формулами, теорема Байеса имеет глубокую интерпретацию. Имеется n взаимоисключающих гипотез $H_1, \ldots, H_n$. Мы знаем их *априорные*[12] вероятности $\mathbb{P}(H_i)$. Мы проводим эксперимент и наблюдаем событие A. Спрашивается: какова *апостериорная*[13] вероятность $\mathbb{P}(H_i|A)$? Теорема Байеса позволяет найти эти апостериорные вероятности, если дополнительно известны вероятности $\mathbb{P}(A|H_i)$ наблюдать событие A при условии верности гипотез H_i.

Пример 3.11. Пусть в задаче из примера 3.9 известно, что вы вытащили белый шар. Какова вероятность того, что вы выбрали первую урну?

Роль гипотез здесь играют события U_1 и U_2, их вероятности нам известны по условию (3.5). Условные вероятности наблюдать событие $White$, состоящее в том, что вы вытянули белый шар, нам известны из (3.6):

$$\mathbb{P}(White|U_1) = 2/3, \quad \mathbb{P}(White|U_2) = 1/2.$$

Тогда, согласно формуле Байеса,

$$\mathbb{P}(U_1|White) = \frac{2/3 \cdot 1/2}{2/3 \cdot 1/2 + 1/2 \cdot 1/2} = 4/7.$$

3.3 Независимость

Независимость — ключевое понятие теории вероятностей, выделяющее ее из абстрактной теории меры: в теории меры такого понятия не существует, потому что отсутствует естественная мотивация его вводить.

Как и прежде, мы живем на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$.

Определение 3.12. События $A, B \in \mathcal{F}$ называются *независимыми*, если

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{3.8}$$

[12] "Априорный" означает "до проведения эксперимента" .

[13] "Апостериорный" значит "после проведения эксперимента" , когда мы получили дополнительные данные — в нашем случае они состоят в том, что произошло событие A.

Легко видеть, что в случае, когда $\mathbb{P}(B) \neq 0$, определение (3.8) эквивалентно соотношению

$$\mathbb{P}(A|B) = \mathbb{P}(A), \tag{3.9}$$

означающему, что от наложения условия B вероятность события A не меняется. Это наблюдение можно рассматривать как мотивировку определения 3.12. Несмотря на более прозрачную интерпретацию формулы (3.9) в сравнении с формулой (3.8), в качестве определения обычно используют последнюю, потому что она удобнее для непосредственной проверки, и в ней отсутствует ограничение $\mathbb{P}(B) \neq 0$. Тут же заметим, что

Задача 3.13. Если $\mathbb{P}(B) = 0$, то событие B независимо с любым событием $A \in \mathcal{F}$.

Сформулируем в виде задачи следующую несложную, но очень полезную лемму, хорошо согласующуюся с нашей интуицией касательно понятия "независимость" .

Задача 3.14. Пусть события $A, B \in \mathcal{F}$ независимы. Докажите, что произвольные события $C \in \{A, A^c\}$ и $D \in \{B, B^c\}$ также независимы.

Определение 3.12 обобщается на случай нескольких событий следующим образом:

Определение 3.15. События $A_1, \dots, A_n$ называются *независимыми в совокупности*, если для всех $2 \leq k \leq n$ и $1 \leq i_1 < \dots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}).$$

При этом говорят, что события $A_1, \dots, A_n$ *попарно независимы*, если A_i и A_j независимы для любых $i \neq j$. Из независимости в совокупности очевидно следует попарная независимость. Обратное неверно:

Задача 3.16. Приведите пример вероятностного пространства $(\Omega, \mathcal{F}, \mathbb{P})$ и трех событий $A, B, C \in \mathcal{F}$, независимых попарно, но не в совокупности.

Подсказка: рассмотрите бросок двух различимых монеток с классическим определением вероятности. Придумайте подходящие события, каждое из которых состоит из двух элементарных исходов.

Когда пишут фразу типа "события $A_1, \dots, A_n$ независимы" , как правило имеют ввиду именно независимость в совокупности, хотя корректно задать уточняющий вопрос.

Понятие независимости в совокупности и попарной независимости естественным образом расширяется на случай счетного числа событий. Точнее, говорят, что события $A_1, A_2, \dots$ независимы в совокупности, если для любого n события $A_1, \dots, A_n$ независимы в совокупности в смысле определения 3.15. Заметим, что в этим случае непрерывность меры влечет

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots) = \prod_{j=1}^{\infty} \mathbb{P}(A_{i_j}).$$

Зависимость и независимость далеко не всегда интуитивно очевидны:

Пример 3.17. В примере 2.3 о подбрасывании двух различимых кубиков с классическим определением вероятности зададимся вопросом о независимости события $A = \{\text{сумма результатов бросков кратна 3}\}$ и $B = \{\text{сумма результатов бросков четна}\}$. Вероятность события A была вычислена в примере 2.9 и составляет $\mathbb{P}(A) = 5/18$. Очевидно, $\mathbb{P}(B) = 1/2$. С другой стороны,

$$\mathbb{P}(A \cap B) = \mathbb{P}\{(1,5), (2,4), (5,1), (4,2), (3,3), (6,6)\} = \frac{1}{6}.$$

Так как $5/18 * 1/2 = 5/36 \neq 1/6$, события A и B зависимы.

Задача 3.18. Покажите, что при броске одного кубика события, аналогичные рассмотренным в примере выше, независимы.

А в других примерах интуиция работает хорошо:

Пример 3.19. Являются ли независимыми различные броски в примере 2.1, 2.8 о подбрасывании симметричной монеты $n \geq 1$ раз (эквивалентно, о подбрасывании n различимых симметричных монет)? Интуиция говорит "да" , но заданный вопрос нужно как-то трансформировать в математический. Следующая интерпретация кажется разумной: являются ли независимыми в совокупности произвольные события $A_1, \ldots, A_n$ вида

$$A_i = \{\omega = (\omega_1 \ldots, \omega_n) \in \Omega : \omega_i \in B_i\}, \quad \text{где} \quad B_i \subset \{"\text{O}" , "\text{P}" \}?$$

Ответ на этот вопрос оказывается положительным, но мы убедимся в этом только в частном случае, проверив независимость событий $A_1 = \{\omega_1 = "\text{O}" \}$ и $A_2 = \{\omega_2 = "\text{P}" \}$, состоящих в том, что при первом броске выпал орел, а при втором – решка. Действительно,

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}\{\omega_1 = "\text{O}" , \ \omega_2 = "\text{P}" , \ (\omega_i)_{i \geq 3} \text{ произвольны}\} = \frac{2^{n-2}}{2^n} = \frac{1}{4},$$

и рассуждая аналогично находим, что $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 1/2$.

Полная же проверка заявленной независимости в совокупности кажется громоздкой и одна мысль о ней удручает. Но оказывается, ее и не нужно проверять, потому что, как мы покажем в следующем параграфе, такая независимость гарантирована самой структурой вероятностного пространства, представляющегося в виде прямого произведения.

3.4 Вероятностная модель для независимых испытаний

Случай конечного числа экспериментов

Пусть у нас имеется n экспериментов, то есть n вероятностных пространств $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$. Тогда мы можем отвечать на вопросы типа "с какой вероятностью в третьем эксперименте произойдет событие $A_3 \in \mathcal{F}_3$?" Ответ дается вероятностью $\mathbb{P}_3(A_3)$. А как нам ответить на вопрос "с какой вероятностью в первом эксперименте произойдет событие A_1, во втором — A_2, и так далее, а в последнем — A_n, если мы считаем, что эксперименты проводятся независимо?" Для этого необходимо построить новое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, на котором можно было бы "поселить" все n экспериментов независимым образом. То есть так, чтобы события B_i вида "в i-ом эксперименте произошло событие $A_i \in \mathcal{F}_i$" должны быть независимы в совокупности.

Предлагается это делать следующим образом. Возьмем в качестве $(\Omega, \mathcal{F}, \mathbb{P})$ прямое произведение вероятностных пространств $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, см. аппендикс B.3. Тогда упомянутые выше события B_i принимают вид $B_i = \Omega_1 \times \cdots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \cdots \times \Omega_n$, или, записывая более подробно,

$$B_i = \{\omega = (\omega_1, \ldots, \omega_n) \in \Omega : \omega_i \in A_i\}, \qquad A_i \in \mathcal{F}_i.$$

Следующая лемма утверждает, что предложенная конструкция дает именно то, что мы и хотели.

Лемма 3.20. *Для произвольных $A_i \in \mathcal{F}_i$, $1 \leq i \leq n$, события $B_1, \ldots, B_n$ независимы в совокупности и $\mathbb{P}(B_i) = \mathbb{P}_i(A_i)$.*

Доказательство. Равенство $\mathbb{P}(B_i) = \mathbb{P}_i(A_i)$ немедленно следует из определения вероятности $\mathbb{P}$:

$$\mathbb{P}(B_i) = \mathbb{P}_1(\Omega_1) \cdots \mathbb{P}_{i-1}(\Omega_{i-1})\mathbb{P}_i(A_i)\mathbb{P}_{i+1}(\Omega_{i+1}) \cdots \mathbb{P}_n(\Omega_n) = \mathbb{P}_i(A_i).$$

Независимость проверяется аналогично:

$$\mathbb{P}(B_{i_1} \cdots B_{i_k}) = \mathbb{P}(\omega : \omega_{i_1} \in A_{i_1}, \ldots, \omega_{i_k} \in A_{i_k}) = \mathbb{P}_{i_1}(A_{i_1}) \cdots \mathbb{P}_{i_n}(A_{i_n}).$$

□

Пример 3.21. Вероятностное пространство, описывающее исход однократного подбрасывания несимметричной монетки с вероятностью выпадения орла равной $0 \le p \le 1$, разумно выбрать в виде $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$, где $\hat{\Omega} = \{0, 1\}$, $\hat{\mathcal{F}} = 2^{\hat{\Omega}}$ и $\hat{\mathbb{P}}\{1\} = p$, $\hat{\mathbb{P}}\{0\} = 1 - p$. Тогда конструкция, предложенная выше, позволяет нам, без всяких дополнительных проверок и вычислений, предъявить вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, описывающее $n \ge 1$ независимых подбрасываний такой монетки. Оно задается прямым произведением n копий вероятностного пространства $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$, так что мы получаем

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \hat{\Omega}\}, \quad \mathcal{F} = \sigma(\hat{\mathcal{F}} \times \cdots \times \hat{\mathcal{F}}) = 2^{\Omega},$$

и

$$\mathbb{P}(\omega) = \hat{\mathbb{P}}(\omega_1) \cdots \hat{\mathbb{P}}(\omega_n) = p^{\sum_{i=1}^n \omega_i}(1-p)^{n - \sum_{i=1}^n \omega_i}.$$

Случай бесконечного числа экспериментов

Ключевым разделом теории вероятностей, которого мы скоро коснемся, являются так называемые предельные теоремы; например, закон больших чисел и Центральная предельная теорема. В частности, они освещают вопрос "можно ли что-нибудь сказать о статистике результатов большого числа независимых экспериментов, зная лишь некоторые характеристики распределений, описывающих эти эксперименты?" Математическая формулировка подобных вопросов требует взятия предела "число экспериментов n стремится к бесконечности" . Для этого нужно, чтобы вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, на котором живут наши n экспериментов, не зависело от n. Это требует построения вероятностного пространства, на которое влезает счетное число независимых экспериментов.

В прошлом разделе мы построили подобное вероятностное пространство для произвольного фиксированного конечного числа экспериментов n, но, к сожалению, напрямую предложенная конструкция на случай счетного числа экспериментов не обобщается. Действительно, можно увидеть, что счетное прямое произведение $\mathcal{F}_1 \times \mathcal{F}_2 \times \ldots$ уже не является полукольцом множеств, так что теорема Каратеодори неприменима (см. аппендикс B.3).

Эта проблема имеет элегантное и естественное решение, которое позволяет строить вероятностное пространство, "поддерживающее" не только лишь счетное, но совершенно любое число экспериментов. Итак, пусть имеется набор вероятностных пространств $(\Omega_t, \mathcal{F}_t, \mathbb{P}_t)_{t \in \mathcal{T}}$, описывающих эксперименты, параметризованные произвольным множеством $\mathcal{T}$. Например, в случае $\mathcal{T} = [0, \infty)$, (непрерывный) параметр t можно интерпретировать как время, и думать, что в каждый момент времени совершается некоторый случайный эксперимент. [14]

Множество элементарных исходов Ω выберем, как и прежде, в виде

$$\Omega = \bigotimes_{t \in \mathcal{T}} \Omega_t.$$

[14]Развитие этой идеи ведет к понятию "случайный процесс" .

Подмножества $C \subset \Omega$ вида

$$C = \{\omega = (\omega_t)_{t \in \mathcal{T}} : \omega_{t_1} \in A_1, \dots, \omega_{t_k} \in A_k\}, \quad \text{где} \quad k \in \mathbb{N}, \quad t_j \in \mathcal{T}, \quad A_j \in \mathcal{F}_{t_j} \quad \forall j \le k, \tag{3.10}$$

будем называть *цилиндрическими*. Менее формально, цилиндрические множества — это такие, которые несут информацию об исходе лишь *конечного* числа экспериментов, при этом не важно каким был исход остальных.

Обозначим символом $\mathcal{C}$ набор всех цилиндрических множеств. Если $|\mathcal{T}| < \infty$, то $\mathcal{C}$ совпадает с прямым произведением сигма-алгебр $\otimes_{t \in \mathcal{T}} \mathcal{F}_t$, однако в общем случае это неверно. Преимущество набора множеств $\mathcal{C}$ перед прямым произведением $\otimes_{t \in \mathcal{T}} \mathcal{F}_t$ видно из следующей задачи.

Задача 3.22. Докажите, что набор всех цилиндрических множеств образует полукольцо.

Определим функцию множеств $\mathbb{P}$ на полукольце $\mathcal{C}$ по формуле

$$\mathbb{P}(C) = \prod_{j=1}^{k} \mathbb{P}_{t_j}(A_j),$$

если $C \in \mathcal{C}$ имеет вид (3.10). Нетрудно видеть, что $\mathbb{P}$ сигма-аддитивна на $\mathcal{C}$. То есть, если непересекающиеся множества $C_1, C_2, \dots \in \mathcal{C}$ удовлетворяют $C := \cup_{i=1}^{\infty} C_i \in \mathcal{C}$, то $\mathbb{P}(C) = \sum_{i=1}^{\infty} \mathbb{P}(C_i)$.

Определение 3.23. Наименьшая σ-алгебра $\sigma(\mathcal{C})$, содержащая всевозможные цилиндрические подмножества, называется *цилиндрической σ-алгеброй*.

Обозначим $\mathcal{F} := \sigma(\mathcal{C})$. По теореме Каратеодори B.13, $\mathbb{P}$ однозначно продолжается до меры на измеримом пространстве $(\Omega, \mathcal{F})$ (которую мы по-прежнему будем обозначать $\mathbb{P}$).

Тройка $(\Omega, \mathcal{F}, \mathbb{P})$ и дает желаемое вероятностное пространство. Действительно, аналогично предыдущему параграфу положим для $s \in \mathcal{T}$

$$B_s = \{\omega = (\omega_t)_{t \in \mathcal{T}} \in \Omega : \omega_s \in A_s\}, \qquad A_s \in \mathcal{F}_s.$$

Лемма 3.24. *Для произвольных $n \ge 2$ и $A_{s_i} \in \mathcal{F}_{s_i}$, $1 \le i \le n$, события $B_{s_1}, \dots, B_{s_n}$ независимы в совокупности и $\mathbb{P}(B_{s_i}) = \mathbb{P}_{s_i}(A_{s_i})$.*

Доказательство повторяет доказательство леммы 3.20.

Хотя, формально говоря, мы и поселили на одно вероятностное пространство произвольное число экспериментов, в конструкцию заложена некоторая априорная счетность – см. определение цилиндрической сигма-алгебры. Последняя лемма вообще рассматривает лишь конечные наборы множеств, но это связано с тем, что определение независимости в совокупности давалось лишь для конечного числа событий, и можно написать аналог леммы и для счетного числа событий. Действительно, рассмотрим событие вида

$$C = \{\omega = (\omega_t)_{t \in \mathcal{T}} \in \Omega : \omega_{s_1} \in A_{s_1},\ \omega_{s_2} \in A_{s_2}, \dots\},$$

для счетного набора событий $A_{s_i} \in \mathcal{F}_i$. Ясно, что оно лежит в цилиндрической сигма-алгебре $\mathcal{F}$, и, используя непрерывность меры, мы находим $\mathbb{P}(C) = \prod_{i=1}^{\infty} \mathbb{P}_{s_i}(A_{s_i})$. Однако, событие

$$C' = \{\omega = (\omega_t)_{t \in \mathcal{T}} \in \Omega : \omega_s \in A_s \quad \text{для всех} \quad s \in \mathcal{T}'\}$$

для множества $\mathcal{T}'$ мощностью континуум, уже не лежит в цилиндрической сигма-алгебре $\mathcal{F}$, так что мы не можем вычислить его вероятность.

3.5 Леммы Бореля-Кантелли

В этом параграфе мы докажем два утверждения, называемые леммами Бореля-Кантелли, отвечающими на вопрос типа "с какой вероятностью в последовательности экспериментов некоторое событие произойдет бесконечное число раз?"

Пусть, как всегда, $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство, а $A_i \in \mathcal{F}$, $i \geq 1$, — последовательность событий на нем. Рассмотрим событие

$$A_\infty = \{\omega \in \Omega : \omega \in A_i \quad \text{для бесконечного числа событий } A_i\} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i.$$

Лемма 3.25 (лемма Бореля-Кантелли)**.** $i)$ *Если* $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$*, то* $\mathbb{P}(A_\infty) = 0$.
$ii)$ *Если* $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ *и события* A_i *независимы в совокупности, то* $\mathbb{P}(A_\infty) = 1$.

Доказательство. $i)$ Используя вложенность событий $B_n := \bigcup_{i=n}^{\infty} A_i$ и непрерывность вероятности, находим

$$\mathbb{P}(A_\infty) = \lim_{n\to\infty} \mathbb{P}(B_n) \leq \lim_{n\to\infty} \sum_{i=n}^{\infty} \mathbb{P}(A_i) = 0,$$

так как ряд $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ сходится.

$ii)$ Заметим, что $A_\infty^c = \bigcup_{n=1}^{\infty} B_n^c$, где $B_n^c = \bigcap_{i=n}^{\infty} A_i^c$. Тогда

$$\mathbb{P}(A_\infty^c) \leq \sum_{n=1}^{\infty} \mathbb{P}(B_n^c).$$

Используя независимость событий A_i^c, находим

$$\mathbb{P}(B_n^c) = \prod_{i=n}^{\infty} \mathbb{P}(A_i^c) = \prod_{i=n}^{\infty} (1 - \mathbb{P}(A_i)) = 0,$$

так как $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$.

□

Используя лемму Бореля-Кантелли, нетрудно показать, что если вы набираете случайный текст, то с вероятностью 1 вы напечатаете Войну и Мир бесконечное число раз.

4 Случайные величины

Эксперименты, для которых применяются вероятностные описания, связаны как правило с числовыми измерениями, и основным объектом для вероятностного описания измерений служит *случайная величина.* Ее можно определить двумя эквивалентными способами:

1. Числовая функция ξ на вероятностном пространстве Ω

2. Структура вероятностного пространства на множестве Ω, состоящем из чисел.

Вторая формулировка вытекает из первой: достаточно рассматривать множество чисел вида $\xi(\omega)$, $\omega \in \Omega$ и сопоставить каждому числу вероятность его прообраза в Ω. Первая же формулировка получается из второй при помощи тождественного отображения $\Omega \to \Omega$. При этом следует помнить, что указанные определения используются в дискретном случае без ограничений, но в более общей ситуации требуют уточнения: например, первая формулировка включает требование измеримости функции. Теперь обсудим все это более подробно.

4.1 Случайные величины и их распределения

Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство.

Определение 4.1. Измеримая функция $\xi : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ называется *случайной величиной.* [15]

Аналогично можно определить случайные величины со значениями в абстрактном измеримом пространстве $(X, \mathcal{G})$ как измеримые функции $\xi : (\Omega, \mathcal{F}) \mapsto (X, \mathcal{G})$. Очевидные примеры: $(X, \mathcal{G}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ и $(X, \mathcal{G}) = (\mathbb{C}^d, \mathcal{B}(\mathbb{C}^d))$. Более сложные примеры возникают в теории случайных процессов, где X часто является каким-либо функциональным пространством (непрерывных функций, соболевским, etc). Чтобы подчеркнуть, что ξ принимает значения в множестве X, иногда говорят, что ξ — X-значная случайная величина.

Однако, в наших лекциях случайные величины почти всегда будут принимать значения либо в $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, либо в $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. В этом последнем случае $\xi = (\xi_1, \ldots, \xi_d)$, и мы будем стараться называть ξ *случайным вектором*, либо d-мерной случайной величиной. Ее компоненты ξ_j при этом являются случайными величинами. Но часто мы будем заговариваться и говорить про ξ просто "случайная величина" . Размерность обычно будет понятна из контекста.

Следующее определение основано на понятии образа меры при отображении, см. аппендикс B.4.

Определение 4.2. Вероятностная мера $\mathbb{P}_\xi := \xi_*(\mathbb{P})$ на измеримом пространстве $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ называется *распределением случайной величины* ξ.

Другими словами, $\mathbb{P}_\xi$ — такая вероятностная мера на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, что

$$\mathbb{P}_\xi(A) = \mathbb{P}\big\{\omega \in \Omega : \xi(\omega) \in A\big\}, \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

Обычно запись $\mathbb{P}\big\{\omega \in \Omega : \xi(\omega) \in A\big\}$ сокращают до $\mathbb{P}\big\{\xi \in A\big\}$ или $\mathbb{P}(\xi \in A\big)$. Мы тоже будем следовать этой славной традиции.

Аналогичным образом определяют распределение случайного вектора $\xi = (\xi_1, \ldots, \xi_d)$ как меру $\mathbb{P}_\xi := \xi_*(\mathbb{P})$ на $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. В этом случае меру $\mathbb{P}_\xi$ также называют *совместным распределением* случайных величин $\xi_1, \ldots, \xi_d$. Аналогичное определение работает и для случайных величин со значениями в абстрактном измеримом пространстве.

Пример 4.3. Рассмотрим эксперимент, состоящий из $n \geq 1$ независимых подбрасываний монетки с вероятностью выпадения орла равной p и вероятностную модель для этого эксперимента, предложенную в примере 3.21. Пусть ξ —случайная величина, описывающая число орлов, выпавших за n бросков,

$$\xi(\omega) := \sum_{i=1}^{d} \omega_i.$$

Тогда ее распределение $\mathbb{P}_\xi$ задано соотношением

$$\mathbb{P}_\xi\{k\} = \mathbb{P}\{\xi = k\} = C_n^k p^k (1-p)^{n-k}, \qquad k = 1, 2, \ldots, n.$$

Такое распределение называется *биномиальным* с параметрами n, p.
Пишут $\xi \sim Binomial(n, p)$.

[15] В определении случайной величины используется борелевская сигма-алгебра, а не более широкая сигма-алгебра множеств, измеримых по Лебегу, чтобы иметь возможность брать композиции случайных величин, не теряя измеримости.

Важное замечание. Когда мы обсуждаем случайные величины, обычно нам интересно отвечать на вопросы типа "с какой вероятностью ξ принимает значения в множестве A?" , или, другими словами, "чему равно $\mathbb{P}_\xi(A)$?" . Поэтому, как правило, единственное, что нам нужно знать о случайной величине, — это ее распределение $\mathbb{P}_\xi$. Конкретная же структура вероятностного пространства $(\Omega, \mathcal{F}, \mathbb{P})$ и явный вид функции $\xi : \Omega \mapsto \mathbb{R}$ бывают важны довольно редко, и обычно при решении задач они не уточняются.

Каждая случайная величина ξ определяет новое вероятностное пространство

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\xi).$$

Зачем же тогда рассматривать двухслойную конструкцию, состоящую из исходного вероятностного пространства $(\Omega, \mathcal{F}, \mathbb{P})$ и пространства $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\xi)$, если можно сразу рассмотреть последнее и получить интересующую нас информацию о вероятностях вида $\mathbb{P}(\xi \in A) = \mathbb{P}_\xi(A)$? Это упростило бы теорию, так как можно было бы не вводить понятие случайной величины. Если нас интересует лишь одна случайная величина ξ, то так сделать действительно можно, но на практике часто оказывается важным иметь возможность ввести на *том же* вероятностном пространстве новую случайную величину ζ с другим интересующим нас распределением (тогда, например, можно будет поточечно вычислять сумму $\xi(\omega) + \zeta(\omega)$). В примере 4.3 это может быть, скажем, результат третьего броска. Вероятностное пространство $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\xi)$ может оказаться слишком бедным для того, чтобы случайная величина ζ была на нем определена. Поэтому предложенная двухслойная конструкция удобна. В ней мы всегда заранее и по умолчанию предполагаем, что исходное вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ достаточно богато, чтобы на нем были определены все рассматриваемые нами случайные величины и используемые конструкции.

Пример 4.4. Пусть $M \in \mathcal{B}(\mathbb{R}^d)$ — ограниченное борелевское множество в $\mathbb{R}^d$. Рассмотрим случайный вектор ξ со значениями в $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ с распределением

$$\mathbb{P}(\xi \in A) = \mathbb{P}_\xi(A) = \frac{Leb(A \cap M)}{Leb(M)}, \quad \forall A \in \mathcal{B}(\mathbb{R}^d),$$

где Leb обозначает меру Лебега. Такое распределение называют *равномерным* распределением на M, а случайный вектор ξ – *равномерно распределенным* на M. Пишут $\xi \sim Uniform(M)$.

Отметим, что, в соответствии с замечанием выше, мы указали лишь распределение случайной величины, но не выбор вероятностного пространства. Его можно выбирать по-разному, и разным выборам будут соответствовать разные случайные величины. Простейший выбор такой:

$$\Omega = \mathbb{R}^d, \quad \mathcal{F} = \mathcal{B}(\mathbb{R}^d), \quad \mathbb{P}(A) = \frac{Leb(A \cap M)}{Leb(M)} \quad \text{и} \quad \xi(\omega) = \omega.$$

Это в точности упомянутое выше вероятностное пространство вида $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_\xi)$.

4.2 Типы случайных величин и векторов

Выделяют три типа случайных величин и векторов, классификация происходит из классификации их распределений, см. аппендикс B.5. А именно, случайную величину (или вектор) ξ называют:

• *Дискретной*, если ее распределение $\mathbb{P}_\xi$ является дискретной мерой. То есть, существует не более, чем счетное множество $X \subset \mathbb{R}^d$, такое что $\mathbb{P}(\xi \in X) = 1$. Это близко к тому, чтобы сказать, что случайная величина принимает не более, чем счетное число значений, но не совсем то. Точки $a \in \mathbb{R}^d$, такие что $\mathbb{P}_\xi(a) = \mathbb{P}\{\xi = a\} > 0$, называют *атомами меры* $\mathbb{P}_\xi$.

• *Сингулярной*, если ее распределение является сингулярной непрерывной мерой. То есть, $\mathbb{P}(\xi = x) = 0$ для всех $x \in \mathbb{R}^d$, но существует множество $X \in \mathcal{B}(\mathbb{R}^d)$ нулевой меры Лебега, такое что $\mathbb{P}(\xi \in X) = 1$. Такие случайные величины довольно экзотичны.

• *абсолютно непрерывной*, если ее распределение абсолютно непрерывно. То есть, случайная величина ξ (а точнее, ее распределение $\mathbb{P}_\xi$) имеет плотность p_ξ — неотрицательную интегрируемую функцию, такую что

$$\mathbb{P}(\xi \in A) = \int_A p_\xi(x)\,dx, \quad \forall A \in \mathcal{B}(\mathbb{R}^d).$$

В частности, $\int_{\mathbb{R}^d} \rho_\xi(x)\,dx = 1$.

Замечание 4.5. Обратите внимание на небольшое несоответствие в терминологии между мерами и случайными величинами: меру называют сингулярной, если она либо дискретна, либо непрерывна сингулярна. А случайную величину называют сингулярной, если ее распределение только лишь непрерывно сингулярно.

Сингулярные и абсолютно непрерывные случайные величины ξ вместе называют *непрерывными*. Их характеризует равенство $\mathbb{P}(\xi = x) = 0$ для всех $x \in \mathbb{R}^n$.

Кроме указанных выше трех типов случайных величин, разумеется, бывают случайные величины смешанных типов. По теореме B.19 о разложении меры распределение произвольной случайной величины ξ однозначно представляется в виде

$$\mathbb{P}_\xi = c_d\nu_d + c_s\nu_s + c_a\nu_a,$$

где ν_d, ν_s и ν_a — дискретная, непрерывная сингулярная и абсолютно непрерывная вероятностные меры, а c_d, c_s, c_a — неотрицательные константы (появляющиеся из нормировки мер μ_d, μ_s, μ_a из теоремы B.19, чтобы получить вероятностные меры $\nu_{d,s,a}$: $\mu_a = c_a\nu_a$, и аналогично для $\nu_{d,s}$).

Замечание 4.6. Аналогично обсуждению ниже примера 4.3, можно построить вероятностное пространство и случайные величины ξ_d, ξ_s, ξ_a на нем, так, чтобы их распределения совпадали с мерами ν_d, ν_s, ν_a. Разумеется, как правило $\xi \neq c_d\xi_d + c_s\xi_s + c_a\xi_a$. Более того, чтобы вычислить распределение суммы случайных величин, нужно знать их совместное распределение.

В дальнейшем нам будет полезно следующее следствие теоремы Фубини.

Лемма 4.7. *Пусть $\xi = (\xi_1, \dots, \xi_d)$ — абсолютно непрерывный случайный вектор с плотностью p_ξ. Тогда его компоненты ξ_j тоже являются абсолютно непрерывными случайными величинами, а их плотность имеет вид*

$$p_{\xi_j}(x) = \int_{\mathbb{R}^{n-1}} p_\xi(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)\,dx_1 \cdots dx_{j-1}dx_{j+1} \cdots dx_n, \qquad x \in \mathbb{R}. \quad (4.1)$$

Доказательство. Для удобства обозначений проведем доказательство для $j = 1$. Для любого $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}(\xi_1 \in A) = \int_{A \times \mathbb{R}^{n-1}} p_\xi(x_1, \dots x_n)\,dx_1 \cdots dx_n = \int_A p_{\xi_1}(x_1)\,dx_1,$$

где в последнем равенстве мы использовали теорему Фубини, а p_{ξ_1} определено в (4.1). $\square$

Задача 4.8. Покажите, что обратное утверждение не верно: из абсолютной непрерывности компонент случайного вектора не следует абсолютная непрерывность самого вектора.

Пример 4.9. Биномиальная случайная величина $\xi \sim Binomial(n, p)$ дискретна. Ее распределение имеет вид $\mathbb{P}_\xi = \sum_{k=0}^{n} \mathbb{P}(\xi = k)\delta_k = C_n^k p^k q^{n-k}\delta_k$, где $q = 1 - p$, а δ_k обозначает дельта-меру в точке k (см. (B.3)). Последнее равенство удобнее записывать в виде таблички:

ξ	0	$\ldots$	k	$\ldots$	n
$\mathbb{P}_\xi$	$\binom{n}{0}q^n$	$\ldots$	$\binom{n}{k}p^k q^{n-k}$	$\ldots$	$\binom{n}{n}p^n$

Пример 4.10. Распределение дискретного случайного вектора (α, β) также удобно записывать в виде таблички. Например,

$\beta \setminus \alpha$	-1	0	$+1$
-1	$1/8$	$1/12$	$1/8$
0	$1/12$	$1/6$	$1/12$
$+1$	$1/8$	$1/12$	$1/8$

В частности, здесь написано, что $\mathbb{P}(\alpha = 0, \beta = 1) = 1/12$.

Пример 4.11. Равномерно распределенная на $M \in \mathcal{B}(\mathbb{R}^d)$ случайная величина $\xi \sim Uniform(M)$ абсолютно непрерывна. Ее плотность имеет вид

$$p_\xi(x) \equiv \frac{1}{|M|}\mathbb{I}_M(x),$$

где $\mathbb{I}_M$ обозначает индикаторную функцию множества M, а $|M|$ — его меру Лебега. Обычно последнее равенство и приводится в качестве определения равномерного распределения. В частности, если $\xi \sim Uniform([0, 1])$, то $p_\xi(x) = \mathbb{I}_{[0,1]}(x)$.

Пример сингулярной случайной величины мы дадим ниже, после того как определим, что такое функции распределения.

4.3 Функции распределения

Обычно самый практичный способ работать с распределением случайной величины — использовать табличку, если случайная величина дискретна, и использовать плотность, если она абсолютно непрерывна. Эти способы не универсальны, так как, помимо экзотических сингулярных случайных величин, бывают еще и случайные величины смешанного типа. Это одна из многих причин, из-за которой удобно ввести следующий универсальный инструмент, однако же хорошо работающий только в размерности $d = 1$.

Итак, почти до конца этого параграфа мы рассматриваем случайную величину ξ со значениями лишь в $\mathbb{R}$.

Определение 4.12. Функция $F_\xi : \mathbb{R} \mapsto [0, 1]$, определенная равенством

$$F_\xi(x) := \mathbb{P}(\xi \le x),$$

называется *функцией распределения случайной величины* ξ.

Эквивалентно, $F_\xi(x) = \mathbb{P}_\xi\big((-\infty, x]\big)$.

Лемма 4.13. *Функция распределения F_ξ однозначно определяет распределение $\mathbb{P}_\xi$. То есть, из равенства $F_\xi = F_\eta$ следует $\mathbb{P}_\xi = \mathbb{P}_\eta$.*

Доказательство. Полуинтервалы вида $(-\infty, x]$ образуют полукольцо множеств. Поэтому утверждение леммы следует из теоремы Каратеодори. $\square$

Перечислим некоторые свойства функции распределения.

Лемма 4.14. *Функция распределения F_ξ случайной величины ξ удовлетворяет следующим свойствам:*
1) F_ξ не убывает, то есть $F_\xi(x) \le F_\xi(y)$ при $x \le y$.
2) $\lim\limits_{x\to-\infty} F_\xi(x) = 0$ и $\lim\limits_{x\to\infty} F_\xi(x) = 1$.
3) F_ξ непрерывна справа, то есть $\lim\limits_{y\to x+} F_\xi(y) = F_\xi(x)$.

Доказательство. Первое свойство следует из того, что событие $\{\omega : \xi(\omega) \le x\}$ содержится в событии $\{\omega : \xi(\omega) \le y\}$ при $x \le y$. Второе и третье свойства следуют из непрерывности меры. Действительно, пересечение событий $\{\omega : \xi(\omega) < x\}$ по всем $x < 0$ пусто, а значит

$$\lim_{x\to-\infty} F_\xi(x) = \lim_{x\to-\infty} \mathbb{P}(\xi \le x) = 0,$$

согласно лемме B.9(4). Так как $F_\xi(x) = 1 - \mathbb{P}(\xi > x)$, предел $\lim\limits_{x\to\infty} F_\xi(x)$ находится аналогично, что влечет пункт 2 леммы. Пункт 3 аналогичным образом следует из леммы B.9(3). $\square$

Итак, к настоящему моменту мы увидели, что функция распределения F_ξ случайной величины ξ однозначно определяет ее распределение $\mathbb{P}_\xi$ и обладает свойствами 1-3, перечисленными в лемме. Оказывается, верно и обратное, в следующем смысле.

Всякую функцию $F : \mathbb{R} \mapsto [0,1]$, удовлетворяющую свойствам 1-3 леммы 4.14, будем называть *функцией распределения.* Рассмотрим функцию множеств μ, определенную на всех полуинтервалах вида $(x, y]$, где $x \ge -\infty$, $y \in \mathbb{R}$, по правилу

$$\mu\big((x,y]\big) = F(y) - F(x), \qquad F(-\infty) := 0. \tag{4.2}$$

Напомним, что множество таких полуинтервалов образует полукольцо. Можно показать, что функция множеств μ — счетно аддитивна, см. [KS, §3.2]. Тогда, по теореме Каратеодори, она единственным образом продолжается до меры μ на минимальной сигма-алгебре, порожденной рассматриваемым полукольцом. А последняя и является борелевской сигма-алгеброй $\mathcal{B}(\mathbb{R})$. Для иллюстрации отметим, что полученная мера удовлетворяет соотношениям

$$\mu([a,b]) = F(b) - \lim_{x\to a-} F(x), \qquad \mu((a,b)) = \lim_{x\to b-} F(x) - F(a).$$

В частности, $\mu(\mathbb{R}) = 1$, так что полученная мера является вероятностной.

Рассмотрим теперь вероятностное пространство $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ и случайную величину на нем $\xi(\omega) \equiv \omega \ \forall \omega \in \mathbb{R}$. Тогда, согласно (4.2),

$$F_\xi(x) = \mu(\xi \le x) = F(x).$$

Таким образом, мы построили случайную величину ξ с функцией распределения F_ξ, равной F. Заметим, что распределение $\mathbb{P}_\xi$ (мера μ выше) определялось однозначно, а вот случайная величина ξ, конечно, не однозначно (например, можно было выбрать какое-нибудь другое вероятностное пространство).

Функции распределения различных типов случайных величин

Рассмотрим случайную величину ξ. Как мы обсудили выше, между распределениями случайных величин $\mathbb{P}_\xi$ на $\mathbb{R}$ (являющимися вероятностными мерами) и функциями распределения F_ξ существует взаимно однозначное соответствие. Следовательно, можно выделить классы функций распределения, соответствующие дискретным, сингулярным непрерывным и абсолютно непрерывным распределениям. Сейчас мы этим и займемся, однако прежде докажем следующую общую лемму.

Лемма 4.15. *Множество точек разрыва функции распределения F_ξ случайной величины ξ не более чем счетно. Оно состоит из всех точек $x \in \mathbb{R}$, в которых $\mathbb{P}(\xi = x) > 0$.*

Доказательство. Первое утверждение — хорошо известное свойство монотонных функций. Для обоснования второго утверждения достаточно рассматривать точки разрыва слева, потому что функция распределения непрерывна справа, согласно лемме 4.14(3),. Тогда желаемое утверждение следует из соотношения

$$\mathbb{P}(\xi = x) = \lim_{\varepsilon \to 0+} \mathbb{P}(x - \varepsilon < \xi \leq x) = F_\xi(x) - \lim_{\varepsilon \to 0+} F_\xi(x - \varepsilon),$$

где в первом равенстве мы использовали непрерывность вероятности. □

- *Дискретные случайные величины*

Лемма 4.16. *Случайная величина ξ дискретна в том и только том случае, когда ее функция распределения F_ξ имеет ступенчатый вид.*

Доказательство. Если случайная величина дискретна, то утверждение очевидно: функция распределения F_ξ имеет ступенчатый вид со скачками в точках $x \in \mathbb{R}$, для которых $\mathbb{P}(\xi = x) > 0$. Высота скачка в точности равна $\mathbb{P}(\xi = x)$.

Обратное следует из того, что функция распределения F_ξ, как и всякая монотонная функция, не может иметь более, чем счетное число скачков. Учитывая ступенчатость F_ξ, отсюда легко видеть, что ξ дискретна. □

- *Абсолютно непрерывные случайные величины*

Напомним, что функция $F(x)$ вида

$$F(x) = \int_{-\infty}^{x} p(y)\, dy, \tag{4.3}$$

где функция p интегрируема, называется *абсолютно непрерывной*. Функция p при этом называется *производной* функции F, что обозначается обычным образом $p = F'$. Такая терминология оправдана: если функция p непрерывна, то, по теореме о дифференцируемости интеграла по верхнему пределу, действительно имеем $p = F'$ в классическом смысле. В общем случае, когда p лишь интегрируема, из курса анализа известно, что функция F дифференцируема почти во всех (по мере Лебега) точках $x \in \mathbb{R}$, и в этих точках верно соотношение $p(x) = F'(x)$.

Лемма 4.17. *Случайная величина ξ абсолютно непрерывна в том и только том случае, когда ее функция распределения F_ξ абсолютно непрерывна. При этом плотность распределения p_ξ является производной функции распределения,*

$$F'_\xi(x) = p_\xi(x). \tag{4.4}$$

Доказательство. Если случайная величина ξ абсолютно непрерывна, то она имеет плотность p_ξ. А значит

$$F_\xi(x) = \mathbb{P}(\xi \leq x) = \int_{-\infty}^{x} p_\xi(y)\, dy, \tag{4.5}$$

то есть функция распределения F_ξ абсолютно непрерывна.

Обратно, допустим, что случайная величина ξ имеет абсолютно непрерывную функцию распределения F_ξ, то есть, существует интегрируемая функция p, для которой верно (4.5). Покажем, что в этом случае p является плотностью распределения ξ, так что ξ — абсолютно непрерывна. Действительно, достаточно проверить, что

$$\mathbb{P}(\xi \in A) = \int_A p(y)\, dy \qquad \forall A \in \mathcal{B}(\mathbb{R}).$$

Ввиду однозначности продолжения меры, утверждаемой теоремой Каратеодори, достаточно проверить последнее равенство для множеств вида $A = (a, b]$. А для них имеем

$$\mathbb{P}((a,b]) = \mathbb{P}(\xi \le b) - \mathbb{P}(\xi \le a) = F_\xi(b) - F_\xi(a) = \int_a^b p(y)\, dy,$$

что и требовалось. □

Пример 4.18. Пусть $\xi \sim Uniform([a,b])$. Тогда

$$F_\xi(x) = \begin{cases} 0, & \text{если } x \le a, \\ \frac{x-a}{b-a}, & \text{если } a < x \le b, \\ 1, & \text{если } x > b. \end{cases}$$

- *Сингулярные случайные величины.*

Согласно лемме 4.15, функция распределения сингулярной случайной величины непрерывна. Однако, она не может быть представлена в виде интеграла от плотности (4.3), а множество ее точек роста имеют нулевую меру Лебега. Мы не будем останавливаться на этом экзотическом случае подробно, а вместо этого рассмотрим пример такой функции распределения: это знаменитая канторова лестница.

Как известно, она непрерывна, множество ее точек роста является канторовым множеством и имеет нулевую меру Лебега. Эта функция почти всюду дифференцируема, но в точках дифференцируемости ее производная равна нулю. При этом в нуле она равна нулю, а в единице — единице, так что для случайной величины ξ с такой функцией распределения имеем $\mathbb{P}(\xi \in [0,1]) = 1$.

Многомерные функции распределения

В многомерном случае функции распределения определяются аналогично, однако с ними гораздо менее удобно работать. Но тем не менее они полезны: например, с их помощью бывает удобно проверять независимость случайных величин, речь об этом пойдет в следующем параграфе.

Пусть $\xi = (\xi_1, \ldots, \xi_d)$ — случайный вектор.

Определение 4.19. Функция $F_\xi(x) : \mathbb{R}^d \mapsto [0,1]$, определенная равенством

$$F_\xi(x) = \mathbb{P}\{\xi_1 \le x_1, \ldots, \xi_d \le x_d\},$$

где $x = (x_1, \ldots, x_d)$, называется *функцией распределения* случайного вектора ξ.

Лемма 4.20. *Функция распределения F_ξ однозначно определяет распределение $\mathbb{P}_\xi$ в смысле леммы 4.13.*

Доказательство этой леммы аналогично доказательству ее одномерного аналога — леммы 4.13. Также имеет место аналог леммы 4.14, однако формулировки в многомерном случае оказываются сложнее, и мы их не приводим. См. [KS, теорема 1.47].

4.4 Независимость случайных величин

Теперь введем следующее ключевое определение.

Определение 4.21. Случайные величины ξ и η, определенные на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$, называются *независимыми*, если

$$\mathbb{P}(\xi \in A, \eta \in B) = \mathbb{P}(\xi \in A)\mathbb{P}(\eta \in B) \quad \forall\, A, B \in \mathcal{B}(\mathbb{R}). \tag{4.6}$$

Так как равенство (4.6) можно переписать в виде

$$\mathbb{P}(\xi^{-1}(A) \cap \eta^{-1}(B)) = \mathbb{P}(\xi^{-1}(A))\mathbb{P}(\eta^{-1}(B)),$$

то определение независимости эквивалентно звучит так: для любых борелевских множеств $A, B \in \mathcal{B}(\mathbb{R})$ события $\xi^{-1}(A)$ и $\eta^{-1}(B)$ независимы.

Третий способ записать определение независимости — следующий:

$$\mathbb{P}_{(\xi,\eta)}(A \times B) = \mathbb{P}_\xi(A)\mathbb{P}_\eta(B) \quad \forall\, A, B \in \mathcal{B}(\mathbb{R}). \tag{4.7}$$

Определение 4.22. Случайные величины $\xi_1, \ldots, \xi_n$ называются *независимыми в совокупности*, если

$$\mathbb{P}(\xi_1 \in A_1, \ldots, \xi_n \in A_n) = \mathbb{P}(\xi_1 \in A_1) \cdots \mathbb{P}(\xi_n \in A_n) \quad \forall\, A_j \in \mathcal{B}(\mathbb{R}). \tag{4.8}$$

Заметим, что, в отличие от определения 3.15 независимости в совокупности событий, в определении 4.22 мы не рассматриваем всевозможные подпоследовательности $i_1, \ldots, i_k$. Их рассмотрение не было бы ошибкой, но это излишне, так как это покрывается выбором в качестве некоторых A_j всего множества Ω.

Определения выше дословно переносятся на случай случайных векторов заменой $\mathbb{R}$ на $\mathbb{R}^d$, а также случайных величин со значениями в абстрактном измеримом пространстве $(X, \mathcal{G})$.

Предложение 4.23. *Независимость в совокупности случайных величин $\xi_1, \ldots, \xi_n$ эквивалентна каждому из следующих тождеств:*
1) $\mathbb{P}_{(\xi_1,\ldots,\xi_n)} = \mathbb{P}_{\xi_1} \times \cdots \times \mathbb{P}_{\xi_n}$;
2) $F_{(\xi_1,\ldots,\xi_n)}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n)$;
3) в дополнительном предположении, что случайный вектор $\xi = (\xi_1, \ldots, \xi_n)$ абсолютно непрерывен:

$$p_{(\xi_1,\ldots,\xi_n)}(x_1, \ldots, x_n) = p_{\xi_1}(x_1) \cdots p_{\xi_n}(x_n).$$

4) в дополнительном предположении, что случайные величины $\xi_1, \ldots, \xi_n$ дискретны:

$$\mathbb{P}(\xi_1 = x_1, \ldots, \xi_n = x_n) = \mathbb{P}(\xi_1 = x_1) \cdots \mathbb{P}(\xi_n = x_n),$$

для всех $x_j \in \operatorname{supp} \mathbb{P}_{\xi_j}$ (то есть таких x_j, что $\mathbb{P}(\xi_j = x_j) > 0$ — атомов мер $\mathbb{P}_{\xi_j}$).

Доказательство. 1) Из формулы 1) независимость следует по ее определению, аналогично (4.7). Обратно, пусть случайные величины независимы в совокупности. То есть,

$$\mathbb{P}_{(\xi_1,\ldots,\xi_n)}(A_1 \times \cdots \times A_n) = \mathbb{P}_{\xi_1}(A_1) \times \cdots \times \mathbb{P}_{\xi_n}(A_n) \quad \forall A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}).$$

Теперь требуемый результат следует из единственности продолжения меры с полукольца множеств $\mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R})$ на минимальную сигма-алгебру, им порожденную (которая совпадает с $\mathcal{B}(\mathbb{R}^n)$), обеспеченной теоремой Каратеодори.

2) Если случайные величины $\xi_1, \dots, \xi_n$ независимы в совокупности, то формула 2) следует по определению. Обратно, пусть верна формула 2). Рассмотрим случайный вектор η, имеющий распределение $\mathbb{P}_\eta = \mathbb{P}_{\xi_1} \times \dots \times \mathbb{P}_{\xi_n}$. Тогда $F_\eta(x) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n)$, а значит, по предположению, $F_\eta = F_{(\xi_1,\dots,\xi_n)}$. Так как, согласно лемме 4.20, функция распределения однозначно определяет распределение, мы находим $\mathbb{P}_\eta = \mathbb{P}_{(\xi_1,\dots,\xi_n)}$, то есть имеет место формула 1). Значит, по уже доказанному, случайные величины $\xi_1, \dots, \xi_n$ независимы.

3) Напомним, что абсолютная непрерывность компонент ξ_j вектора ξ следует из леммы 4.7. Пусть случайные величины $\xi_1, \dots, \xi_n$ независимы. Тогда для любых событий $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ из формулы (4.8) следует

$$\int_{A_1\times\dots\times A_n} p_{(\xi_1,\dots,\xi_n)}(x_1,\dots,x_n)\,dx_1\dots dx_n = \int_{A_1} p_{\xi_1}(x_1)\,dx_1 \cdots \int_{A_n} p_{\xi_n}(x_n)\,dx_n.$$

Это влечет равенство 3). Обратно, пусть имеет место равенство 3). Тогда из него сразу следует равенство 2), которое по уже доказанному влечет желаемую независимость.

4) Независимость по определению влечет требуемое равенство. Докажем обратное: для любых $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned}
\mathbb{P}\{\xi_1 \in A_1, \dots, \xi_n \in A_n\} &= \sum_{x_1 \in A_1 \cap \operatorname{supp} \mathbb{P}_{\xi_1}} \cdots \sum_{x_n \in A_n \cap \operatorname{supp} \mathbb{P}_{\xi_n}} \mathbb{P}(\xi_1 = x_1, \dots, \xi_n = x_n) \\
&= \sum_{x_1 \in A_1 \cap \operatorname{supp} \mathbb{P}_{\xi_1}} \mathbb{P}(\xi_1 = x_1) \cdots \sum_{x_n \in A_n \cap \operatorname{supp} \mathbb{P}_{\xi_n}} \mathbb{P}(\xi_n = x_n) \\
&= \mathbb{P}\{\xi_1 \in A_1\} \cdots \mathbb{P}\{\xi_n \in A_n\}.
\end{aligned}$$

□

Следствие 4.24. *Если известны распределения $\mathbb{P}_{\xi_1}, \dots, \mathbb{P}_{\xi_n}$ каждой из случайных величин $\xi_1, \dots, \xi_n$, и известно, что случайные величины независимы, то известно и совместное распределение этих случайных величин, то есть распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$:*

$$\mathbb{P}_\xi = \mathbb{P}_{\xi_1} \times \dots \times \mathbb{P}_{\xi_n}.$$

Задача 4.25. Пусть в примере 4.4 с равномерным распределением $M = [0,1]^d$. Докажите, что компоненты случайного вектора ξ независимы и имеют равномерное распределение на $[0,1]$. Верно ли это утверждение, если M — шар в $\mathbb{R}^d$?

Задача 4.26. Пусть вероятностное пространство имеет вид $([0,1], \mathcal{B}([0,1]), Leb)$. Рассмотрим пару функций $x(t) = 2t$, $y(t) = 1 - t^2$ как случайные величины. Зависимы ли они? Тот же вопрос для пары отображений $x(t) = \operatorname{sign}[\sin(2\pi t)]$ и $y(t) = \operatorname{sign}[\sin(4\pi t)]$.

Возвращаясь к связи интуитивного и формального понимания независимости, посмотрим, как обстоит дело с традиционно понимаемой функциональной зависимостью; обычно это утверждение $\xi_2 = g(\xi_1)$ о возможности выразить одну функцию через другую.

Задача 4.27. 1) Пусть случайные величины ξ_1, ξ_2 удовлетворяют соотношению $\xi_2 = g(\xi_1)$ для некоторой функции g. Верно ли, что они зависимы? Если нет, то при каких условиях?

2) Найдите пример двух зависимых случайных величин, таких что не существует функции g, для которой соотношение выше было бы верным.

Таким образом, зависимость в смысле теории вероятностей отвечает более широкому смыслу, нежели функциональная зависимость.

Сумма независимых случайных величин

Напомним, что для интегрируемых функций $f, g \in L_1(\mathbb{R})$ определена их свертка

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)\,dy;$$

она также удовлетворяет $f * g \in L_1(\mathbb{R})$.

Лемма 4.28. *Если случайные величины ξ и η абсолютно непрерывны и имеют плотности p_ξ и p_η, то случайная величина $\xi+\eta$ также абсолютно непрерывна и*

$$p_{\xi+\eta} = p_\xi * p_\eta.$$

Доказательство. Согласно предложению 4.23(3), случайный вектор (ξ, η) абсолютно непрерывен и имеет плотность $p_{(\xi,\eta)}(x,y) = p_\xi(x)p_\eta(y)$. Вычислим с ее помощью функцию распределения случайной величины $\xi+\eta$:

$$F_{\xi+\eta}(t) = \int_{x+y\le t} p_{(\xi,\eta)}(x,y)\,dxdy = \int_{-\infty}^{t}\int_{-\infty}^{\infty} p_{(\xi,\eta)}(s-y,y)\,dyds = \int_{-\infty}^{t} (p_\xi * p_\eta)(s)\,ds,$$

где мы использовали замену переменных $(s,y) := (x+y, y)$. □

Модель независимых случайных величин

Пусть имеется последовательность вероятностных пространств $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, $1 \le i \le n$, на i-ом пространстве определена случайная величина ξ_i. Как «поселить» их все на одно вероятностное пространство независимым образом? Для этого можно воспользоваться конструкцией, предложенной в параграфе 3.4. Действительно, достаточно выбрать вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ в виде прямого произведения вероятностных пространств $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, и рассмотреть случайный вектор $\widetilde{\xi} = (\widetilde{\xi}_1, \dots, \widetilde{\xi}_n)$ с компонентами $\widetilde{\xi}_i(\omega) := \xi_i(\omega_i)$. Нетрудно видеть, что $\mathbb{P}_{\widetilde{\xi}_i} = \mathbb{P}_{\xi_i}$ и компоненты вектора $\widetilde{\xi}$ независимы в совокупности. Убедимся в этом: в

$$\mathbb{P}\{\widetilde{\xi}_j \in A\} = \mathbb{P}\{\omega \in \Omega : \xi_j(\omega_j) \in A\} = \mathbb{P}_j\{\xi_j \in A\} = \mathbb{P}_{\xi_j}(A), \qquad \forall A \in \mathcal{B}(\mathbb{R}).$$

Далее, для всех $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_{\widetilde{\xi}}(A_1 \times \dots \times A_n) = \mathbb{P}\{\omega : \xi_1(\omega_1) \in A_1, \dots, \xi_n(\omega_n) \in A_n\} = \mathbb{P}_{\xi_1}(A_1)\cdots\mathbb{P}_{\xi_n}(A_n) = \mathbb{P}_{\widetilde{\xi}_1}(A_1)\cdots\mathbb{P}_{\widetilde{\xi}_n}(A_n).$$

В случае бесконечного числа экспериментов нужно аналогичным образом воспользоваться соответствующей конструкцией вероятностного пространства из параграфа 3.4.

5 Моменты

5.1 Матожидание

Пусть как всегда $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство, а $\xi : \Omega \mapsto \mathbb{R}$ — случайная величина.

Определение 5.1. *Математическим ожиданием* $\mathbb{E}\xi$ случайной величины ξ называется интеграл

$$\mathbb{E}\xi = \int_\Omega \xi(\omega)\,d\mathbb{P}(\omega).$$

Математическим ожиданием n-мерного случайного вектора $\eta = (\eta_1, \ldots, \eta_n)$ называют вектор матожиданий его компонент $\mathbb{E}\eta := (\mathbb{E}\eta_1, \ldots, \mathbb{E}\eta_n)$. См. параграф B.6, где напоминается конструкция интеграла Лебега и его основные свойства.

Как видно из определения, матожидание существует не для любой случайной величины. Те, для которых оно существует, образуют банахово пространство $L_1(\Omega, \mathbb{P})$ с нормой $\|\xi\| := \mathbb{E}|\xi|$.

Пример 5.2. Пусть Ω — не более, чем счетное множество. Тогда функция ξ автоматически является простой и, согласно опрделению интеграла Лебега от простых функций,

$$\mathbb{E}\xi = \sum_{\omega \in \Omega} \xi(\omega) \mathbb{P}\{\omega\}. \tag{5.1}$$

В частности, если $|\Omega| = n$ и вероятность определена классическим образом $\mathbb{P}(\omega) = 1/n$, матожидание случайной величины ξ совпадает со средним арифметическим ее значений

$$\mathbb{E}\xi = \sum_{\omega \in \Omega} \frac{\xi(\omega)}{n}.$$

Из определение выше создается иллюзорное впечатление, что матожидание $\mathbb{E}\xi$ зависит от структуры вероятностного пространства, а не только от распределения $\mathbb{P}_\xi$. Это не так: согласно теореме B.24 о замене переменной в интеграле Лебега, имеет место следующий результат.

Теорема 5.3. *Пусть $f : \mathbb{R}^n \mapsto \mathbb{R}$ — измеримая функция, а $\eta = (\eta_1, \ldots, \eta_n)$ — случайный вектор. Тогда*

$$\mathbb{E}f(\eta) = \int_{\mathbb{R}^n} f(x)\, d\mathbb{P}_\eta(x), \quad \text{где} \quad x = (x_1, \ldots, x_n).$$

Интеграл в правой части равенства существует тогда и только тогда, когда существует интеграл в левой части.

В частности,

$$\mathbb{E}\xi = \int_{\mathbb{R}} x\, d\mathbb{P}_\xi(x).$$

Замечание 5.4. Можно показать, что интеграл Лебега по мере $\mathbb{P}_\xi$ совпадает с интегралом Стилтьеса по dF_ξ, где F_ξ — функция распределения ξ. Поэтому в литературе вместо $d\mathbb{P}_\xi$ часто пишут dF_ξ.

Пример 5.5. Пусть ξ — дискретная случайная величина, а $X \subset \mathbb{R}$ — не более, чем счетное множество, такое что $\mathbb{P}\{\xi \in X\} = 1$. Тогда

$$\mathbb{E}\xi = \int_{\mathbb{R}} x\, d\mathbb{P}_\xi(x) = \sum_{x \in X} x\, \mathbb{P}_\xi\{x\}. \tag{5.2}$$

Последняя формула допускает наглядную интерпретацию: если думать про x как про координату точки, а про $\mathbb{P}_\xi\{x\}$ — как про массу, сидящую в этой точке, то $\mathbb{E}\xi$ дает центр масс системы точек X. Формула (5.1) допускает аналогичную интерпретацию, где ω — имена точек, $\xi(\omega)$ — их координаты, а $\mathbb{P}_\xi\{\omega\}$ — веса, ассоциированные с точками. Заметим, что при этом координаты различных точек могут совпадать. Если же собрать вместе все точки с одинаковыми координатами и сложить их массы, то формула (5.1) превращается в точности в формулу (5.2). Проверьте это!

Пример 5.6. Если случайная величина ξ абсолютно непрерывна, то $d\mathbb{P}_\xi(x) = p_\xi(x)\,dx$, так что

$$\mathbb{E}\xi = \int_{-\infty}^{\infty} x p_\xi(x)\,dx.$$

Пример 5.7. • Если $\xi \sim Bernoulli(p)$, $0 \le p \le 1$, то есть $\mathbb{P}(\xi = 1) = p$ и $\mathbb{P}(\xi = 0) = 1 - p$, то

$$\mathbb{E}\xi = 0 \times (1-p) + 1 \times p = p.$$

- Если $\xi \sim Exp(\lambda)$, $\lambda > 0$, то есть $p_\xi(x) = \lambda e^{-\lambda x}\mathbb{I}_{[0,\infty)}(x)$, то

$$\mathbb{E}\xi = \lambda \int_0^{\infty} x e^{-\lambda x}\,dx = -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx = \lambda^{-1}.$$

Здесь пропущен параграф или 2

6 Важнейшие распределения

6.1 Дискретный случай

В естествознании имеется базовый набор примеров дискретных вероятностных пространств и связанных с ними случайных величин. Значение этих общепринятых примеров чрезвычайно велико, без их знания невозможно никакое дальнейшее углубление в теорию, поскольку более сложные конструкции обычно принято сводить к базовым. Очень подробно свойства этих базовых моделей разобраны в книге В.Феллер "Введение в теорию вероятностей и ее приложения", том 1, поиск на указатель терминов.

В этом параграфе всегда $0 \le p \le 1$ и $q = 1 - p$.

6.2 Распределние Бернулли (индикатор)

Случайная величина принимает всего два значения:

ξ	0	1
$\mathbb{P}_\xi$	q	$p.$

Говорят, что ξ имеет распределение Бернулли, пишут $\xi \sim Bernoulli(p)$. Еще иногда говорят, что ξ— индикаторная случайная величина

Модель

Распределение возникает при рассмотрении опыта с двумя исходами (однократное бросание неровной монеты), которые мы назовем герб и решка или успех и неудача.

Производящая функция

Очевидно, имеет вид $Q(z) = (q + pz)$.

Основные характеристики

Непосредственным вычислением устанавливаем, что $\mathbb{E}\eta = p$ и $\operatorname{Var}\eta = pq$.

6.3 Биномиальное распределение

Следующее распределение случайной величины ξ называется *биномиальным распределением* с параметрами $n \geq 1$ и p:

ξ	0	$\ldots$	k	$\ldots$	n
$\mathbb{P}_\xi$	$\binom{n}{0}q^n$	$\ldots$	$\binom{n}{k}p^k q^{n-k}$	$\ldots$	$\binom{n}{n}p^n$

Пишут $\xi \sim Binomial(n, p)$ или $B(n, p)$.

Модель

Биномиальное распределение возникает при рассмотрении n-кратных независимых повторений опыта с двумя исходами. Таким образом, про ξ можно думать как про число орлов, выпавших при n независимых бросаниях монеты с вероятностью выпадения орла равной p. Точнее, нетрудно проверить, что случайная величина

$$S_n := \xi_1 + \cdots + \xi_n, \tag{6.1}$$

где $\xi_j \sim Bernoulli(p)$ и независимы, имеет биномиальное распределение $B(n, p)$.

Производящая функция

В силу (6.1), производящая функция биномиального распределения равна $(q + pz)^n$.

Основные характеристики

Воспользуемся свойствами математического ожидания и дисперсии, и формулой (6.1), чтобы убедиться в том, что $\mathbb{E}(\xi) = np$ (аддитивность математического ожидания), $\mathrm{Var}(\xi) = npq$ (дисперсия суммы независимых с.в. равна сумме их дисперсий).

6.4 Геометрическое распределение

Это (не путать с геометрическими вероятностями!) распределение целочисленной неотрицательной случайной величины, принимающей бесконечно много значений:

α	0	1	$\ldots$	k	$\ldots$
$\mathbb{P}_\alpha$	p	qp	$\ldots$	$q^k p$	$\ldots$

Пишут $\xi \sim Geom(p)$.

Модель

Геометрическое распределение возникает при рассмотрение независимых повторений эксперимента с двумя исходами до наступления первого успеха, где вероятность успеха в единичном эксперименте равна p. Если речь идет о подбрасывании монеты (успех – выпадения герба), то элементарные исходы кодируются последовательностями букв вида: Г, РГ, РРГ, РРРГ,. . . . Вероятности элементарных событий выбираются по правилу независимых повторений. На исходе РРРРРР. . . РГ (k решек подряд) случайная величина α принимает значение равное k (число неудач до первого успеха). Надо только проверить, что это действительно распределение, в самом деле:

$$p + qp + q^2 p + \ldots = p(1 + q + q^2 + \ldots) = \frac{p}{1-q} = 1$$

Производящая функция

$$Q(z) = p + qpz + q^2pz^2 + \ldots = p(1 + qz + (qz)^2 + \ldots) = \frac{p}{1 - qz}$$

Основные характеристики

По формуле $\mathbb{E}(\alpha) = Q'_\alpha(1) = \frac{q}{p}$

- Найдите самостоятельно дисперсию геометрического распределения.

6.5 Распределение Паскаля

β	0	1	$\ldots$	k	$\ldots$
	p^r	$\binom{r}{1}qp^r$	$\ldots$	$\binom{r+k-1}{k}q^kp^r$	$\ldots$

, где $r \geq 1$. Пишут $\beta \sim BN(n, p)$ (Blaise Pascal).

Проверка свойств и заодно обобщенная формула Бинома Ньютона

Для распределения Паскаля часто используется также название «отрицательное биномиальное распределение», его происхождение объясняется так:

$$\binom{x}{k} = \frac{x(x-1)(x-2)\ldots(x-k+1)}{k!} \Longrightarrow \qquad \forall r > 0 \qquad \binom{-r}{k} = (-1)^k\binom{r+k-1}{k}$$

Следовательно вероятности можно записать в ином, более удобном для запоминания, виде:

β	0	1	$\ldots$	k	$\ldots$
	p^r	$\binom{-r}{1}(-q)p^r$	$\ldots$	$\binom{-r}{k}(-q)^kp^r$	$\ldots$

Общеизвестна каноническая формула Бинома Ньютона, которая доказывается раскрытием скобок и использованием числа сочетаний для подсчета способов получить при приведении подобных членов данную степень. В частности:

$$(1-q)^r = \sum_{k=0}^{r}\binom{r}{k}(-q)^k$$

Рассмотрим теперь степенной ряд $1 + q + q^2 + \ldots = 1/(1-q)$ и продифференциируем его почленно r раз, в результате получится ряд

$$\frac{(-1)^r r!}{(1-q)^{r+1}} = r! + \frac{(r+1)!}{1!}q + \frac{(r+2)!}{2!}q^2 + \ldots = r!\sum_{k=0}^{\infty}\binom{r+k}{k}q^k$$

$$(-1)^{r-1}p^{-r} = \frac{(-1)^{r-1}}{(1-q)^r} = (-1)^{r-1}\sum_{k=0}^{\infty}\binom{r+k-1}{k}q^k = (-1)^{r-1}\sum_{k=0}^{\infty}(-1)^k\binom{-r}{k}(-1)^k(-q)^k$$

$$(1-q)^{-r} = \sum_{k=0}^{\infty}\binom{-r}{k}(-q)^k$$

Последнее равенство – это обобщение Бинома Ньютона для отрицательных степеней, а предпоследняя объясняет, почему сумма всех вероятностей в распределении Паскаля действительно равна 1.

Модель

Распределение Паскаля возникает при рассмотрение независимых повторений эксперимента с двумя исходами до наступления r-го успеха. Если речь идет о подбрасывании монеты (успех – выпадения герба), то элементарные исходы кодируются заканчивающимися буквой Г последовательностями вида: РРРГРРГРГ. . . РГ, общее число букв Г в последовательности равно r. Если число решек в последовательности равно k, то случайная величина β принимает на таком элементарном исходе значение k (число неудач до r-го успеха). Вероятности элементарных событий выбираются по правилу независимых повторений. Нетрудно заметить, что всего таких последовательностей фиксированной длины $r+k$ существует в точности $\binom{r+k-1}{r-1} = \binom{r+k-1}{k}$.

Заметим, что неудачи располагались между последовательными успехами – пусть α_1 – число неудач до первого успеха, α_2 – число неудач от первого успеха до второго и так далее. Мы видим, что $\beta = \alpha_1 + \alpha_2 + \ldots + \alpha_r$. Теперь осталось заметить, что случайные величины α_i взаимно независимы и каждая из них имеет геометрическое распределение.

Производящая функция

Согласно свойству производящей функции для суммы независимых случайных величин производящая функция распределения Паскаля совпадает с

$$\left(\frac{p}{1-qz}\right)^r$$

Основные характеристики

Представление $\beta = \sum \alpha_k$ сразу дает ответ для математического ожидания распределения Паскаля: $\mathbb{E}(\xi) = rq/p$. Дисперсия суммы независимых случайных величин равна сумме дисперсий слагаемых, выразите ее в явном виде самостоятельно.

6.6 Формула Пуассона

Рассмотрим теперь случай перехода к пределу в биномиальном распределении при одновременном увеличении $n \to \infty$ и уменьшении $p(n)$: $p(n) = \lambda/n + o(1/n)$. В этом случае $P\left(S(n) = m\right) \to \frac{\lambda^m}{m!}\exp(-\lambda)$. Действительно:

$$P\left(S(n) = m\right) = \binom{n}{m} p^m(n)\left(1-p(n)\right)^{n-m} = \frac{n(n-1)\ldots(n-m+1)}{m!}\left(\frac{\lambda}{n} + o(1/n)\right)^m \left(1 - \frac{\lambda}{n} - o(1/n)\right)^{n-m}$$

Теперь заметим, что $(\lambda/n + o(1/n))^m = (\lambda/n)^m + o(1/n^m)$ и далее

$$\lim_{n\to\infty} n\ldots(n-m+1)\,(\lambda/n)^m = \lambda^m \quad \text{а также} \lim_{n\to\infty}\left(1 - \frac{\lambda}{n} - o(1/n)\right)^{n-m} = e^{-\lambda}(\text{второй замечательный предел}$$

Это вычисление взято из книги В.Тутубалин "Теория вероятностей и случайных процессов. Основы математического аппарата и прикладные аспекты", Глава 1, раздел 4.3, в котором стоит посмотреть любопытные примеры интерпретаций формулы Пуассона на практике.

6.7 Распределение Пуассона,его производящая функция и характеристики

Это дискретное неотрицательное целочисленное распределение, принимающее значения k с вероятностью $\frac{\lambda^k}{k!}e^{-\lambda}$.

γ	0	$\dots$	k	$\dots$
	$e^{-\lambda}$	$\dots$	$\frac{\lambda^k}{k!}e^{-\lambda}$	$\dots$

Вспоминая разложение экспоненты в ряд нетрудно видеть, что производящая функция пуассоновского распределения с параметром λ равна $\exp(\lambda(z-1))$ (проверить!). Заметим, что из этого вытекает, что сумма двух *независимых* пуассоновских величин с параметрами λ и μ будет опять пуассоновской случайной величиной уже с параметром $\lambda+\mu$. Зная явный вид производящей функции несложно проверить, что математическое ожидание пуассоновского распределения равно его дисперсии и равно λ.

Практическая интерпретация распределения Пуассона

Типичное применение – оценка вероятности возникновения нескольких маловероятных событий. Например, известно, что вероятность дефекта у типовой детали равна $p \ll 1$, тогда вероятность обнаружить в партии из m деталей менее k дефектных оценивается из формулы $p_0+p_1+\ldots+p_{k-1}$, где $p_i = \frac{(mp)^j}{j!}e^{-mp}$. При этом используется соображение, что математическое ожидание числа дефектных деталей в партии из m деталей равно mp – соответствующему параметру распределения Пуассона, совпадающему как указано выше с его математическим ожиданием..

Другой пример: нам сообщают, что вероятность неисправности самолета равна 10^{-6} за один час полета, то в принципе ясно что такое число могло возникнуть как частное $p = n_T/T$ от деления общего числа зафиксированных неисправностей n_T на большое количество летных часов T. Тогда, поскольку доставшийся нам параметр $\mu = 10^{-6}$ весьма мал и относится к схеме «повторений опыта каждый час» разумно считать, что для наблюдений за время $t \ll T$ часов имеет место распределение Пуассона с параметром $\lambda = \mu t$, так что можно оценивать вероятности возникновения, скажем, более трех неисправностей за время t (что может служить основой для чрезвычайной ситуации). Важным приложением также является учет вероятности того, что за время $t \ll T$ часов не возникнет ни одной неисправности, это, естественно, вычисляется по формуле $e^{-\mu t}$, что обычно понимается как *вероятность того, что время ожидания неисправности превосходит* t. Такая модель для появления редких событий во времени имеет наименование *пуассоновского потока событий с параметром* μ.

A Напоминание: комбинаторика и простейшие формулы

A.1 Простейшие базовые формулы

- **Формула включений и исключений**. Пусть $|A|$ обозначает число элементов в конечном множестве A. Тогда для (конечного) семейства конечных множеств A_i, $i = 1, 2, \ldots n$ выполнено:

$$|\bigcup_{i=1}^{n} A_i| = \sum_{i}^{n}|A_i| - \sum_{i_1<i_2}|A_{i_1}\cap A_{i_2}| + \sum_{i_1<i_2<i_3}|A_{i_1}\cap A_{i_2}\cap A_{i_3}| - \ldots + (-1)^{n-1}|A_{i_1}\cap A_1\cap A_2\cap\ldots\cap A_n|$$

  Доказательство математической индукцией по подсчету вхождений каждого элемента из $\bigcup_{i=1}^{n} A_i$ в множества из правой части равенства.

- **Таблицы свойств**: пусть имеется k списков, в каждом таком списке перечислено N_k характеристик, надо сосчитать сколько возможных вариантов набора характеристик (по одной из каждого списка). Например, если списков два, то варианты удобно представлять в виде позиций в прямоугольной таблице – это дает ответ $N_1 \cdot N_2$. Многомерный аналог этого подхода дает ответ $N_1 \cdot N_2 \cdot \ldots N_k$.

- **Размещения**: сколько способов построить n солдат в шеренгу длины k. Для правого фланга есть n вариантов, следующее по порядку место заполняется $n-1$ способами итд. Всего вариантов получается $n \cdot (n-1) \cdot (n-2) \cdot \ldots (n-k+1)$, что традиционно обозначается как A_n^k.

- **Перестановки** Размещения с $n = k$ традиционно называются перестановками, обозначаются $\prod_n = n!$, где по определению $0! = 1$.

- **Сочетания**: сколько способов собрать из n солдат взвод в k человек. Ясно, что такой взвод можно построить в шеренгу $\prod_k$ способами, откуда вытекает число способов формирования взвода $\frac{A_n^k}{\prod_k} = \frac{n!}{k!(n-k)!}$, оно называется числом сочетаний или биномиальным коэффициентом, обозначалось ранее как C_n^k, но в современной литературе [16] как $\binom{n}{k}$.

A.2 Некоторые примеры

Перестановки разнотипных объектов или перестановки с повторениями

Пусть имеются предметы k различных типов. Сколько различных перестановок можно сделать в последовательности из n_1 предметов первого типа, n_2 предметов второго типа, $\ldots$, n_k предметов k -го типа? Число элементов в последовательности равно $n = \sum_{i=1}^{k} n_k$. Поскольку различимы элементы только по их типам, то общее число перестановок будет меньше $\prod_n$: некоторые перестановки надо отождествить. В самом деле, однотипные предметы можно переставлять между собой и это даст неотличимую от исходной последовательность. Для разных типов это можно делать независимо и таких «внутренних» перестановок будет соответственно $n_1!$ в первом типе, $n_2!$ во втором итд. Таким образом общее число действительно различных перестановок получится таким:

$$\frac{(n_1 + n_2 + \ldots + n_k)!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

A.2.1 Разложения шаров по ящикам. Различимые и неразличимые объекты

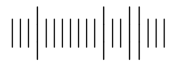
Пусть имеются n шаров и $k \leqslant n$ ящиков, мы собираемся сосчитать количество конфигураций, когда все шары разложены по ящикам. Совершенно очевидно, что прежде всего необходимо договориться, какие кофигурации следует считать разными, это приводит к следующщим задачам с разными условиями:

1. **Все ящики различимы между собой и все шары различимы между собой**. Условие означает, что можно пронумеровать предметы и тогда для каждого шара возникает ровно k возможностей, по формуле таблиц общее число раскладок будет k^n

2. **Все ящики различимы между собой, а все шары неразличимы между собой**. Можно закодировать каждую такую раскладку схематической картинкой:

[16]кроме того для любого действительного x полагают $\binom{x}{k} = \frac{x(x-1)(x-2)\ldots(x-k+1)}{k!}$

палочки, обозначающие стенки плотно стоящих ящиков слева направо и шары между ними. Достаточно укзать только $k-1$ стенку – самая левая и самая правая стенки ничего не прибавляют к знанию о раскладке. А теперь, если шары тоже нарисовать палочками, но поменьше размером, то получится такая, например, картинка, кодирующая раскладку 3,7,2,0,4 шестнадцати одинаковых шаров по пяти различимым (по их порядку) ящикам:



В общем случае всего палочек $n+k-1$ и из них $k-1$ длинных. Всего таких картинок (и раскладок!) получается $\binom{n+k-1}{k-1}$.

3. **Все ящики неразличимы между собой и все шары неразличимы между собой**. Здесь все кодируется числом способов разбиения числа n на не более, чем $k \leqslant n$ ненулевых слагаемых, порядок которых неважен. Поэтому всякому такому разбиению $n = n_1 + n_2 + \ldots + n_j$ можно сопоставить картинку: $j \leqslant k$ выровненных по левой границе горизонтальных полосок (каждая из некоторого числа клеточек, всего клеточек n), нарисованных друг под другом в порядке (нестрогого) убывания длин [17]. Такая картинка см. Рис.2 из клетчатых полосок называется диаграммой Юнга. Нас интересует подсчет всех таких табличек с общим числом клеток n и числом строк не более k.

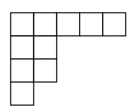


Рис. 2: Пример диаграммы Юнга

Первым делом заметим, что каждую такую табличку можно *транспонировать* – строки по порядку выписать столбцами – тогда опять получится диаграмма Юнга из n клеток уже с любым (естественно, не превосходящим n) количеством строк, но у которой длина любой строки теперь не превосходит k. Мы свели исходную задачу к подсчету $p(n,k)$ – количества разбиений числа n на слагаемые, *каждое из которых* не превосходит k. Продолжение вычислений с диаграммами Юнга см. в разделе методов A.3.1.

4. **Все ящики неразличимы между собой, а все шары различимы между собой**. То есть речь идет о подсчете количества неупорядоченных разбиений множества $\{1, 2, \ldots n\}$ на не более чем k подмножеств. Количество разбиений $\{1, 2, \ldots n\}$ *в точности на i непустых множеств* может быть явно указано, оно называется *числом Стирлинга второго рода* и обычно обозначается как $\left\{ n \atop i \right\}$. Поэтому в нашей задаче про разложение различимых шаров в неразличимые ящики окончательный ответ дается суммой чисел Стирлинга $\left\{ n \atop i \right\}$ по i от единицы до k. Явное вычисление чисел Стирлинга см. далее разделе методов A.3.1.

[17] французские математики предпочитают рисовать их по возрастанию

A.3 Методы вычислений

A.3.1 Рекуррентные соотношения. Треугольник Паскаля

Биномиальные коэффициенты имеют много тождеств и рекуррентных соотношений, например:

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

Это соотношение порождает возможность вычислять биномиальные коэффициенты последовательно выписывая значения, выдаваемые рекуррентным соотношением, в треугольную таблицу (треугольник Паскаля), строки которой принято нумеровать от нуля.

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1 \\
. & . & . & . & . & . & . & . & .
\end{array}$$

Из треугольника Паскаля возможно вычислить и множество других очень полезных комбинаторных формул. Индукцией по номеру строки несложно проверить, что в треугольнике Паскаля k стоят коэффициенты многочлена $(a+b)^n$ последовательно возрастающие как раз в соответствии с рекуррентным соотношением. В частности, при $a = b = 1$ возникает тождество:

$$\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n} = 2^n$$

В докомпьютерную эпоху последовательные вычисления при больших n, r биномиальных коэффициентов $\binom{n}{r}$ даже и с помощью треугольника Паскаля представляли трудность, поэтому возникли асимптотические формулы для факториалов, дающие приблизительные вычисления по формуле $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Часто встречается асимптотическая формула Стирлинга, объяснение которой приведено в разделе про биномиальное распределение.

Производящие функции

В математике часто встречается следующий прием для вычислений. Пусть $a_0, a_1, a_2, \ldots$ — произвольная числовая последовательность $\{a_n\}$. Формальный степенной ряд $\sum_{n=0}^{\infty} a_n q^n$ от переменной q называется производящей функцией для этой последовательности. Например, для последовательности из одних единиц ее производящая функция геометрическая прогрессия — формальный ряд для $\frac{1}{1-q}$, а для (конечной) последовательности биномиальных коэффициентов $\binom{n}{0}, \binom{n}{1} \ldots \binom{n}{n}$ производящая функция есть $(1+q)^n$ итп. Явное знание вида производящей функции позволяет дифференциированием находить ее ряд Маклорена, коэффициенты которого дадут исходную последовательность. Мы применим производящие функции к поставленным выше задачам о разложениях шаров в ящики, больше примеров найдется, например, в книге С.Ландо "Лекции по комбинаторике".

Диаграммы Юнга. Подсчет числа $p(n, k)$ для разбиений

Мы сосредоточимся на вычислении производящей функции $P_k(q)$ для последовательности $\{p(n, k)\}$ с фиксированным k. Ясно что $P_1(q) = \frac{1}{1-q}$ потому что каждое число единственным образом разбивается в сумму единиц. Теперь заметим, что количество способов разбить число n в сумму слагаемых, *каждое из которых* равно двум — это либо 1, если n четно, либо 0, если n нечетно. Чередование единиц и нулей отвечает производящей функции $\frac{1}{1-q^2}$, а $P_2(q) = \frac{P_1(q)}{1-q^2}$. Действительно, раскроем скобки, но не будем пока приводить подобные члены в произведении рядов

$$(1 + q + q^2 + q^3 + \ldots)(1 + q^2 + q^4 + q^6 + \ldots)$$

Каждое слагаемое после раскрытия скобок имеет вид $q^r q^{2s}$ и каждому такому слагаемому можно сопоставить разбиение числа $r+2s$ в сумму r единиц и s двоек. А после приведения подобных членов коэффициент при q^n окажется как раз $p(n, 2)$. Рассуждая аналогичным образом получаем, что производящая функция $P_k(q) = \sum_n p(n,k)q^n$ равна

$$\frac{1}{(1-q)(1-q^2)\ldots(1-q^k)} = \prod_{m=1}^{k} (1-q^m)^{-1}$$

Последовательным дифференциированием правой части мы найдем ее ряд Маклорена и тем самым необходимый коэффициент $p(n, k)$ — достаточно сложный путь!

Вычисление чисел Стирлинга второго рода

Рассмотрим Y – все *эпиморфные* отображения f множества $S_1 = \{1, 2, \ldots n\}$ на множество $S_2 = \{1, 2, \ldots k\}$. Каждое такое отображение делит S_1 в точности на k кусков P_i так, что $f(P_i) = i$. Ясно, что поскольку порядок кусков нам не важен, то искомых отображений будет $k!\left\{ n \atop k \right\} = |Y|$. С другой стороны по формуле таблиц все (то есть уже не обязательно эпиморфные) такиие функции составляют множество X из в точности k^n элементов, обозначим множество тех из них, которые в образе не содержат j через X_j. Понятно, что $Y = \bigcap_{j=1}^{k}(X \setminus X_j)$, а потому

$$|Y| = |X - \bigcup_{j=1}^{k} X_j| = k^n - |\bigcup_{j=1}^{k} X_j|$$

При этом ясно, что $|X_j| = (k-1)^n$ и опять-таки по формуле таблиц

$$\sum_{1 \leqslant i_1 < i_2 < \ldots < i_j \leqslant k} |X_{i_1} \cap X_{i_2} \cap \ldots X_{i_j}| = \binom{k}{j}(k-j)^n$$

Применяя формулу включений и исключений получаем

$$k!\left\{ n \atop k \right\} = |Y| = k^n - \left[\sum_{j=1}^{k}(-1)^{j+1}\binom{k}{j}(k-j)^n\right] = \sum_{j=0}^{k}(-1)^j\binom{k}{j}(k-j)^n$$

откуда уже и получается итоговая формула

$$\left\{ n \atop k \right\} = |Y| = \frac{1}{k!}\sum_{j=0}^{k}(-1)^j\binom{k}{j}(k-j)^n$$

A.4 Асимптотики и формула Стирлинга

Начертим график функции $y = \ln x$ и заметим, что величина $I_n = \ln 2 + \ln 3 + \ldots + \frac{1}{2}\ln n$ выражает формулу трапеций для интегральной суммы, оценивающей площадь под графиком функции $y = \ln x$ на отрезке $[1, n]$. Эта интегральная сумма по построению меньше площади под кривой на отрезке $[1, n]$. С другой стороны, для $1 < k < n$ число $\ln k$ в точности равно площади трапеции, образованной осью абцисс, вертикальными прямыми $x = k - \frac{1}{2}$, $x = k + \frac{1}{2}$ и зажатым меж ними отрезком касательной к графику в точке $(k, \ln k)$ – сделайте рисунок всех упомянутых фигур сами! Поэтому

$$\ln\Big[(n-1)!\Big] = \sum_{k=1}^{n-1} \ln k > \int\limits_{[\frac{3}{2}, n-\frac{1}{2}]} \ln x \; dx \quad \text{и} \quad \int\limits_{[n-\frac{1}{2}, n]} \ln x \; dx < \frac{\ln n}{2} \Longrightarrow I_n > \int\limits_{[\frac{3}{2}, n]} \ln x \; dx$$

Имеем неравенства с определенными интегралами от логарифмической функции, которые вычисляются явно так как $\int \ln x \; dx = x \ln x - x$,

$$\int\limits_{[\frac{3}{2}, n]} \ln x \; dx < I_n < \int\limits_{[1, n]} \ln x \; dx$$

$$\left(n + \frac{1}{2}\right) \ln n - n + \frac{3}{2}\left(1 - \ln\frac{3}{2}\right) < \ln n! < \left(n + \frac{1}{2}\right)\ln n - n + 1$$

$$n^{\left(n+\frac{1}{2}\right)} e^{-n} \cdot \beta < n! < n^{\left(n+\frac{1}{2}\right)} e^{-n} \cdot e$$

$$\beta < \frac{n!}{n^{\left(n+\frac{1}{2}\right)} e^{-n}} < e$$

для $\beta = \exp\left(\frac{3}{2}\left(1 - \ln\frac{3}{2}\right)\right) \simeq 1.23586$. То есть мы указали выражение того же порядка величины, что и $n!$, ибо их отношение лежит между 1.23 и 2.7183. Несколько более детальные рассуждения показывают (см. например В.Феллер "Введение в теорию вероятностей и ее приложения том 1, откуда и взято приводимое выше рассуждение), что это отношение с ростом n быстро стремится к пределу, равному $\sqrt{2\pi} \simeq 2.507$ – возникает формула Стирлинга для асимптотики значения факториала. С ее помощью, например, выясняется, что $\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}}$.

B Базовые объекты и конструкции теории меры

В этом аппендиксе мы напоминаем основны теории меры. Доказательств мы не приводим. Зафиксируем некоторое множество Ω.

B.1 Алгебры и сигма-алгебры

Определение B.1. Набор подмножеств $\mathcal{A}$ множества Ω называется *алгеброй*, если
1) $\Omega \in \mathcal{A}$,
2) $A \in \mathcal{A} \quad \Rightarrow \quad A^c := \Omega \setminus A \in \mathcal{A}$,
3) $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cup A_2 \in \mathcal{A}$.

Лемма B.2. *Пусть $\mathcal{A}$ — алгебра. Тогда*
1) $\emptyset \in \Omega$,
2) $A_1, \ldots, A_n \in \mathcal{A} \quad \Rightarrow \quad \cup_{i=1}^n A_i \in \mathcal{A}$, $\cap_{i=1}^n A_i \in \mathcal{A}$,
3) $A, B \in \mathcal{A} \quad \Rightarrow \quad A \setminus B \in \mathcal{A}$.

Пример B.3. Следующие наборы подмножеств Ω образуют алгебру:

1) $\{\emptyset, \Omega\}$,

2) $2^\Omega :=$ множество всех подмножеств Ω.

3) В случае $\Omega = \mathbb{R}$, набор всевозможных конечных объединений множеств вида (a, b), $(a, b]$, $[a, b]$, $[a, b)$, где $a, b \in \mathbb{R}$.

4) Набор всех конечных подмножеств Ω.

5) Набор всех не более, чем счетных подмножеств Ω.

Определение B.4. Набор подмножеств $\mathcal{F}$ множества Ω называется *σ-алгеброй*, если $\mathcal{F}$ — алгебра, замкнутая относительно операции счетного объединения. То есть, если для любых $A_i \in \mathcal{F}$, $i \geq 1$, выполнено $\cup_{i=1}^\infty A_i \in \mathcal{F}$. Пару $(\Omega, \mathcal{F})$, где $\mathcal{F}$ —σ-алгебра, называют *измеримым пространством.*

Лемма B.5. *Если в определении B.4 $\cup_{i=1}^\infty$ заменить на $\cap_{i=1}^\infty$, то получится эквивалентное определение σ-алгебры.*

Алгебры из примеров B.3(1,2,5) являются σ-алгебрами, а из примеров B.3(3,4) —- вообще говоря нет (если в примере B.3(4) предположить, что множество Ω бесконечно).

Можно показать, что для всякого набора $\mathcal{C}$ подмножеств Ω существует единственная наименьшая сигма-алгебра $\mathcal{F}$ на Ω, содержащая $\mathcal{C}$. Это значит, что если $\widetilde{\mathcal{F}}$ — другая σ-алгебра, содержащая $\mathcal{C}$, то она содержит и $\mathcal{F}$. То есть, $\forall A \in \mathcal{F}$ удовлетворяет $A \in \widetilde{\mathcal{F}}$. Наименьшая сигма-алгебра $\mathcal{F}$ обозначается $\sigma(\mathcal{C})$.

Определение B.6. Если Ω — топологическое пространство, то *борелевской* сигма-алгеброй $\mathcal{B}(\Omega)$ называется наименьшая сигма-алгебра, содержащая все открытые подмножества Ω. Эквивалентно, содержащая все замкнутые подмножества. Подмножества $A \in \mathcal{B}(\Omega)$ называются *борелевскими.*

Например, $\mathcal{B}(\mathbb{R}) = \sigma$(все интервалы) $= \sigma$(все множества из примера B.3(3)).

B.2 Мера

Напомним, что функция множеств μ называется *аддитивной*, если для любых не пересекающихся множеств $A_1, \ldots, A_n$ из ее области определения, $A_i \cap A_j = \emptyset$ при $i \neq j$, выполнено $\mu(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$. Разумеется, выполнение последнего равенства для $n = 2$ влечет его выполнение для произвольного n.

Определение B.7. Пусть дано измеримое пространство $(\Omega, \mathcal{F})$. Аддитивная функция множеств $\mu : \mathcal{F} \mapsto [0, \infty)$ называется *мерой* (конечной, неотрицательной), если она обладает свойством счетной аддитивности. То есть, для любых не пересекающихся подмножеств $A_i \in \mathcal{F}$, $i \geq 1$, $A_i \cap A_j = \emptyset$ при $i \neq j$, выполнено

$$\mu\Big(\bigcup_{i=1}^\infty A_i\Big) = \sum_{i=1}^\infty \mu(A_i).$$

Для краткости, говоря "мера" мы обычно мы будем подразумевать конечную и неотрицательную меру. Отметим, что в некоторых источниках σ-аддитивность не включается в определение меры, откуда появляется выражение "σ-аддитивная мера".

Борелевские сигма-алгебры являются самыми типичными областями определения мер.

Пример B.8. 1) Мера Лебега на измеримом пространстве $\big([0,1], \mathcal{B}[0,1]\big)$ — мера.

2) Пусть Ω — не более, чем счетное множество, а функция $p : \Omega \mapsto [0, \infty)$ удовлетворяет соотношению $\sum_{\omega\in\Omega} p(\omega) < \infty$. Тогда функция множеств μ, определенная равенством

$$\mu(A) := \sum_{\omega\in\Omega} p(\omega), \quad A \subset \Omega,$$

задает меру на измеримом пространстве $(\Omega, 2^\Omega)$.

3) Пусть множество Ω — счетно, а $\mathcal{A}$ — алгебра всех конечных подмножеств в Ω и их дополнений. Рассмотрим функцию множеств m, определенную равенствами

$$m(A) = 0, \quad m(A^c) = 1, \quad \text{для любого конечного подмножества } A. \tag{B.1}$$

Эта функция аддитивна, но не счетно-аддитивно, так что она не является мерой. Да и определена она не на сигма-алгебре, а лишь на алгебре.

4) Пусть множество Ω — несчетно, а $\mathcal{A}$ — алгебра всех счетных подмножеств в Ω и их дополнений. Заметим, что $\mathcal{A}$ является σ-алгеброй. Рассмотрим функцию множеств m, определенную соотношением (B.1), где A — счетные подмножества. Она счетно-аддитивна, а значит задает меру.

Важную роль в теории вероятностей играет свойство мер, называемое *непрерывностью* — см. пункты (2-4) следующей леммы.

Лемма B.9. *Пусть $(\Omega, \mathcal{F})$ — измеримое пространство, а $\mu : \mathcal{F} \mapsto [0, \infty)$ — аддитивная функция множеств. Следующие свойства эквивалентны:*
1) [Счетная аддитивность] *μ — счетно-аддитивна, то есть является мерой.*
2) [Непрерывность снизу] *Для любой возрастающей последовательности множеств A_i, $i \ge 1$ (то есть такой, что $A_i \subset A_{i+1}$) выполнено $\mu(A_i) \to \mu(A)$ при $i \to \infty$, где $A := \cup_{i=1}^\infty A_i$. Это можно записать так:*

$$A_i \uparrow A \quad \Rightarrow \quad \mu(A_i) \to \mu(A) \quad \textit{при} \quad i \to \infty.$$

3) [Непрерывность сверху] *Для любой убывающей последовательности множеств B_i, $i \ge 1$ (то есть такой, что $B_i \supset B_{i+1}$), выполнено $\mu(B_i) \to \mu(B)$ при $i \to \infty$, где $B := \cap_{i=1}^\infty B_i$. Это можно записать так:*

$$B_i \downarrow B \quad \Rightarrow \quad \mu(B_i) \to \mu(B) \quad \textit{при} \quad i \to \infty.$$

4) [Непрерывность в нуле] *Для любой убывающей к нулю последовательности множеств C_i, $i \ge 1$ (то есть такой, что $C_i \supset C_{i+1}$ и $\cap_{i=1}^\infty C_i = \emptyset$), выполнено $\mu(C_i) \to 0$ при $i \to \infty$. Это можно записать так:*

$$C_i \downarrow \emptyset \quad \Rightarrow \quad \mu(C_i) \to 0 \quad \textit{при} \quad i \to \infty.$$

Задача B.10. Докажите лемму.

B.3 Прямое произведение пространств с мерой

Пусть $(\Omega_i, \mathcal{F}_i)$ — измеримые пространства с мерами μ_i, $1 \le i \le n$.

Определение B.11. *Прямым произведением* измеримых пространств $(\Omega_i, \mathcal{F}_i)$ называется измеримое пространство $(\Omega, \mathcal{F})$, где $\Omega = \Omega_1 \times \cdots \times \Omega_n$, а $\mathcal{F} = \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_n)$. Здесь $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ обозначает набор множеств вида вида $A_1 \times \cdots \times A_n$ с $A_i \in \mathcal{F}_i$.

Чтобы пояснить это определение, отметим, что прямое произведение $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ сигма-алгебр $\mathcal{F}_i$ вообще говоря уже не является σ-алгеброй. Чтобы это увидеть, достаточно положить $n = 2$, $\Omega_i = \mathbb{R}$ и $\mathcal{F}_i = \mathcal{B}(\mathbb{R})$, и заметить, что разность прямоугольника со вложенным в него меньшим прямоугольником уже не прямоугольник.

Рассмотрим функцию множеств $\mu = \mu_1 \times \cdots \times \mu_n$, определенную на наборе множеств $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ равенством

$$\mu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n), \qquad \forall A_i \in \mathcal{F}_i.$$

Возникает естественный вопрос, можно ли ее продолжить до меры на измеримом пространстве $(\Omega, \mathcal{F})$? Оказывается, можно.

Определение B.12. Набор множеств $\mathcal{R}$ называется *полукольцом*, если $\emptyset \in \mathcal{R}$,

(1) $A \cap B \in \mathcal{R}$ для всех $A, B \in \mathcal{R}$,

(2) Если $A, B \in \mathcal{R}$ и $A \subset B$, то существуют непересекающиеся множества $C_1, \ldots, C_k \in \mathcal{R}$, такие что $B \setminus A = C_1 \cup \cdots \cup C_k$.

Легко проверить, что $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ является полукольцом (удобно представить себе разность вложенных прямоугольников на плоскости). А теперь сформулируем знаменитую

Теорема B.13 (теорема Каратеодори о продолжении меры)**.** *Пусть счетно-аддитивная функция множеств $\nu : \mathcal{R} \mapsto [0, \infty)$ определена на полукольце $\mathcal{R}$. Тогда она однозначно продолжается до меры на минимальной сигма-алгебре $\sigma(\mathcal{R})$, содержащей $\mathcal{R}$.*

Ниже, в лемме B.16, мы проверяем счетную аддитивность функции множеств μ. Тогда, применяя теорему Каратеодори, мы видим, что μ однозначно продолжается до меры на измеримом пространстве $(\Omega, \mathcal{F})$, которую мы будет обозначать тем же символом μ.

Определение B.14. Измеримое пространство $(\Omega, \mathcal{F})$, снабженное мерой μ, называется *прямым произведением* измеримых пространств $(\Omega_i, \mathcal{F}_i)$, снабженных мерами μ_i.

Пример B.15. Пусть $\Omega_i = [a_i, b_i]$, с действительными $a_i < b_i$, $\mathcal{F}_i = \mathcal{B}([a_i, b_i])$ — борелевская σ-алгебра, а μ_i — мера Лебега на $\mathcal{B}([a_i, b_i])$. Тогда Ω — прямоугольник в $\mathbb{R}^n$, и можно доказать, что $\mathcal{F} = \mathcal{B}(\Omega)$, а μ — мера Лебега на $\mathcal{B}(\Omega)$.

Отметим, что если меры μ_i были вероятностными, то и мера μ также окажется вероятностной:

$$\mu(\Omega) = \prod_{i=1}^{n} \mu_i(\Omega_i) = 1.$$

Лемма B.16. *Функция μ, определенная на полукольце множеств $\mathcal{R} := \mathcal{F}_1 \times \cdots \times \mathcal{F}_n$, σ-аддитивна. То есть, для любого набора не пересекающихся множеств $A_i \in \mathcal{R}$, таких что $\cup_{i=1}^{\infty} A_i \in \mathcal{R}$, выполнено*

$$\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

Доказательство. Проведем доказательство для $n = 2$. Пусть $A_i = B_i \times C_i$, где $B_i \in \mathcal{F}_1$, $C_i \in \mathcal{F}_2$, и $\cup_{i=1}^{\infty} B_i \times C_i = B \times C$, где $B \in \mathcal{F}_1$, $C \in \mathcal{F}_2$. Рассмотрим последовательность функций

$$f_k : \Omega_1 \mapsto \mathbb{R}, \quad f_k(\omega_1) = \sum_{i=1}^{k} \mathbb{I}_{B_i}(\omega_1) \mu_2(C_i),$$

где $\mathbb{I}_{B_i}$ обозначает индикаторную функцию множества B_i. Аналогично положим $f(\omega_1) = \mathbb{I}_B(\omega_1)\mu_2(C)$. Заметим, что $f_k \le \mu_2(B)$ и $f_k(\omega_1) \to f(\omega_1)$ при $k \to \infty$. Тогда, согласно теореме Лебега о мажорируемой сходимости,

$$\mu\big(\cup_{i=1}^k B_i \times C_i\big) = \sum_{i=1}^k \mu(B_i)\mu(C_i) = \int_{\Omega_1} f_k(\omega_1)\, d\mu_1(\omega_1) \overset{k\to\infty}{\to} \int_{\Omega_1} f(\omega_1)\, d\mu_1(\omega_1)$$
$$= \mu_1(B)\mu_2(C) = \mu(B \times C).$$

Доказательство для произвольного n можно провести по индукции. □

B.4 Образ меры при отображении

Пусть имеется два измеримых пространства $(X, \mathcal{F})$ и $(Y, \mathcal{G})$, и на пространстве $(X, \mathcal{F})$ живет мера μ. Пусть $f : (X, \mathcal{F}) \mapsto (Y, \mathcal{G})$ — измеримое отображение. Тогда на пространстве $(Y, \mathcal{G})$ определена мера $f_*(\mu)$, называемая *образом меры μ при отображении f* (pushforward меры), заданная соотношением

$$f_*(\mu)(A) := \mu\big(f^{-1}(A)\big), \quad \forall A \in \mathcal{G}. \tag{B.2}$$

B.5 Типы мер

B.5.1 Сингулярные меры

Определение B.17. Мера μ на $\mathcal{B}(\mathbb{R}^n)$ называется *сингулярной*, если существует множество $A \in \mathcal{B}(\mathbb{R}^n)$ нулевой меры Лебега, такое что $\mu(A) = \mu(\mathbb{R}^n)$.

Другими словами, мера μ сосредоточена на множестве нулевой меры Лебега. Выделяют два типа сингулярных мер: дискретные и непрерывные.

• **Дискретные меры.** Это меры μ на $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, сосредоточенные на не более, чем счетных подмножествах $\mathbb{R}^n$. То есть, существует не более, чем счетное множество $X = \{x_1, x_2, \dots\} \subset \mathbb{R}^n$ для которого $\mu(X) = \mu(\mathbb{R}^n)$.

Напомним, что *дельта-мерой* в точке $x \in \mathbb{R}^n$ называется (вероятностная) мера δ_x, такая что

$$\delta_x(A) = \begin{cases} 1, & \text{если } x \in A, \\ 0, & \text{если } x \notin A, \end{cases} \quad \text{для всех} \quad A \in \mathcal{B}(\mathbb{R}^n). \tag{B.3}$$

Определение выше означает, что дискретная мера μ представляется в виде

$$\mu = \sum_j p_j \delta_{x_j}, \tag{B.4}$$

для некоторых чисел $p_j \ge 0$, таких что $\sum_j p_j < \infty$. Говоря более подробно, (B.4) означает, что

$$\mu(A) = \sum_j p_j \delta_{x_j}(A) = \sum_{j:\, x_j \in A} p_j, \quad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

Мера μ оказывается вероятностной, если $\sum_j p_j = 1$.

В случае $n = 1$ дискретные меры удобно записывать в виде таблиц с двумя строками, в перовой из которых находятся точки x_j, а во второй числа p_j.

• **Непрерывные сингулярные меры.** Это такие сингулярные меры μ на $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, что $\mu(x) = 0$ для любого $x \in \mathbb{R}^n$. Такие меры существуют, однако это экзотика. Для примера см. конец параграфа 4.3, про сингулярные случайные величины.

Нетрудно показать, что любая сингулярная мера μ единственным образом представляется в виде суммы $\mu_d + \mu_s$ дискретной и сингулярной непрерывной мер.

B.5.2 Абсолютно непрерывные меры

Определение B.18. Напомним, что мера μ на $\mathcal{B}(\mathbb{R}^n)$ называется *абсолютно непрерывной*, если для любого множества $A \in \mathcal{B}(\mathbb{R}^n)$ нулевой меры Лебега выполнено $\mu(A) = 0$.

Теорема Радона-Никодима утверждает, что всякая абсолютно непрерывная мера μ имеет *плотность*: существует интегрируемая по Лебегу функция $\rho : \mathbb{R}^n \mapsto [0, \infty)$, такая что

$$\mu(A) = \int_A \rho(x)\, dx, \quad \forall A \in \mathcal{B}(\mathbb{R}^n). \tag{B.5}$$

Это принято обозначать следующим образом:

$$\mu(dx) = \rho(x)\, dx \quad \text{или} \quad d\mu = \rho\, dx.$$

Пусть теперь, наоборот, дана интегрируемая по Лебегу функция $\rho : \mathbb{R}^n \mapsto [0, \infty)$. Из сигма-аддитивности интеграла Лебега следует, что функция множеств μ, определенная соотношением (B.5), является мерой, а значит ρ является ее плотностью. Таким образом, имеется взаимно-однозначное соответствие между интегрируемыми по Лебегу неотрицательными функциями (плотностями) и мерами (с точностью до эквивалентности, обычной для пространств Лебега). Отметим, что, если мера μ — вероятностная, то ее плотность ρ удовлетворяет

$$\int_{\mathbb{R}^n} \rho(x)\, dx = 1.$$

B.5.3 Теорема о разложении меры

Оказывается, тремя типами мер, перечисленными выше, все и исчерпывается.

Теорема B.19. *Для любой меры μ на $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ существуют и единственны дискретная, сингулярная непрерывная и абсолютно непрерывная меры μ_d, μ_s и μ_a, такие что*

$$\mu = \mu_d + \mu_s + \mu_a.$$

Доказательство можно найти в книге [KS, Теорема 3.16]. Оно дано в случае $n = 1$, но легко видеть, что оно работает и для произвольного $n \geq 1$.

B.6 Интеграл Лебега

Пусть $(\Omega, \mathcal{F})$ — измеримое пространство, μ — мера на нем, а $f : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ — измеримая функция. Цель этого параграфа — напомнить как строится интеграл

$$\int_\Omega f(\omega)\, d\mu(\omega)$$

и некоторые его свойства. См. [KS, параграф 3.1] за более подробным изложением.

B.6.1 Построение интеграла

Шаг 1. Пусть сперва функция f принимает не более, чем счетное число значений, то есть

$$f(x) = \sum_{k=1}^{\infty} c_k \mathbb{I}_{A_k}(x), \quad A_k \in \mathcal{F}, \quad c_k \in \mathbb{R}.$$

Будем называть такую функцию измеримой *простой* (легко видеть, что она действительно измерима). [18] Допустим, что $f \geq 0$. Положим

$$\int_\Omega f(\omega)\, d\mu(\omega) := \sum_{k=1}^{\infty} c_k \mu(A_k), \tag{B.6}$$

если ряд в правой части сходится. Если он расходится, положим $\int_\Omega f\, d\mu := \infty$.

Из определения (B.6) сразу же следует

Лемма B.20. *Если f_1, f_2 — простые неотрицательные функции, то*

1. $\displaystyle\int_\Omega f_1\, d\mu \geq \int_\Omega f_2\, d\mu, \quad$ *если* $f_1 \geq f_2$.
2. $\displaystyle\int_\Omega (a_1 f_1 + a_2 f_2)\, d\mu = a_1 \int_\Omega f_1\, d\mu + a_2 \int_\Omega f_2\, d\mu \quad$ *для произвольных констант* a_1, a_2.

Заметим, что утверждения леммы верны и в случае расходящихся интегралов (напомним, что мы пока рассматриваем только неотрицательные функции).

Шаг 2. Обратимся теперь к случаю, когда $f \geq 0$ — произвольная неотрицательная измеримая функция.

Лемма B.21. *Для произвольной измеримой функции $g : \Omega \mapsto \mathbb{R}$ существует последовательность $(g_n)_{n\geq 1}$ простых измеримых функций, таких что $g_n \leq g_{n+1} \leq \cdots \leq g$ и*

$$|g(\omega) - g_n(\omega)| \leq 2^{-n} \qquad \forall \omega \in \Omega.$$

В частности, $g_n \rightrightarrows g$ при $n \to \infty$. Кроме того, если $g \geq 0$, то и $g_n \geq 0$.

Доказательство. Положите

$$g_n(\omega) := \frac{k}{2^n}, \quad \text{если} \quad \frac{k}{2^n} \leq g(\omega) \leq \frac{k+1}{2^n},$$

и пробегите все $k \geq 0$. □

Рассмотрим произвольную последовательность $(f_n)_{n\geq 1}$ неотрицательных простых измеримых функций, таких что $f_n \leq f_{n+1} \leq \cdots \leq f$ и $f_n \rightrightarrows f$ при $n \to \infty$. Тогда, согласно лемме B.20(1), последовательность $\int f_n\, d\mu$ не убывает, а значит сходится к пределу, который мы и назовем интегралом функции f:

$$\int_\Omega f\, d\mu := \lim_{n\to\infty} \int_\Omega f_n\, d\mu. \tag{B.7}$$

Конечно, чтобы такое определение интеграла было корректным, необходима следующая

Теорема B.22. *Предел (B.7) не зависит от выбора последовательности f_n.*

Доказательство этого результата давалось в курсе теории меры. См. также [KS, Теорема 3.4].

Шаг 3. Остается избавиться от условия неотрицательности функции f. Для произвольной измеримой функции f, функции $f_+ := f\mathbb{I}_{f>0}$ и $f_- := -f\mathbb{I}_{f<0}$ — неотрицательны и измеримы, и

$$f = f_+ - f_-.$$

Если интегралы $\int f_+\, d\mu$ и $\int f_-\, d\mu$ конечны, то положим

$$\int_\Omega f\, d\mu = \int_\Omega f_+\, d\mu - \int_\Omega f_-\, d\mu.$$

В этом случае говорят, что функция f интегрируема и пишут $f \in L_1(\Omega, \mu)$.

[18] В литературе простыми часто называют функции, принимающие лишь конечное число значений.

B.6.2 Свойства интеграла Лебега

Если же функция f не интегрируема, причем интеграл $\int f_+ \, d\mu = \infty$, а $\int f_- \, d\mu < \infty$, то говорят, что $\int f \, d\mu = \infty$. Если, наоборот, $\int f_+ \, d\mu < \infty$, а $\int f_- \, d\mu = \infty$, то $\int f \, d\mu = -\infty$. Наконец, если оба интеграла бесконечны, то интеграл $\int f \, d\mu$ не определен.

Переходом к пределу легко проверяется следующий результат:

Лемма B.23. *Утверждения леммы B.20 верны для произвольных измеримых функций f_1, f_2. При этом в пункте (2) достаточно, чтобы конечен был хотя бы один интеграл из правой части.*

Так как $|f| = f_+ + f_-$, конечность интеграла $\int f \, d\mu$ эквивалентна конечности интеграла $\int |f| \, d\mu$, то есть *интегрируемость и абсолютная интегрируемость для интеграла Лебега эквивалентны.* Действительно, если интегралы $\int f_\pm \, d\mu$ конечны, то, согласно лемме B.23, $\int (f_+ + f_-) \, d\mu$ тоже конечен. Обратно, если $\int |f| \, d\mu$ конечен, то, так как $-|f| \le f \le |f|$, опять согласно лемме B.23 интеграл $\int |f| \, d\mu$ конечен.

Пусть $A \in \mathcal{F}$. Как определить интеграл $\int_A f \, d\mu$? На данный момент мы это сделали только для $A = \Omega$. Первый способ — заменить множество Ω множеством A, и повторить построение выше. Другой способ такой:

$$\int_A f \, d\mu := \int_\Omega f \mathbb{I}_A \, d\mu.$$

Легко увидеть, что оба способа дают один и тот же результат. Отметим важное свойство интеграла Лебега — счетную аддитивность. А именно, если $A = \cup_{i=1}^\infty A_i$, $A_i \in \mathcal{F}$ — не пересекающиеся множества, то

$$\int_A f \, d\mu = \sum_{i=1}^\infty \int_{A_i} f \, d\mu$$

для любой измеримой функции f, для которой конечен интеграл из левой части.

B.6.3 Замена переменной в интеграле Лебега

Пусть имеется второе измеримое пространство $(X, \mathcal{G})$ и измеримая функция $h : (\Omega, \mathcal{F}) \mapsto (X, \mathcal{G})$. Напомним, что на $(X, \mathcal{G})$ определена мера $h_*\mu$, перенесенная с помощью отображения h, см. (B.2).

Теорема B.24. *Для любой измеримой функции $f : X \mapsto \mathbb{R}$,*

$$\int\limits_\Omega f(h(\omega)) \, d\mu(\omega) = \int\limits_X f(x) \, df_*\mu(x).$$

Интеграл в правой части определен в том и только том случае, когда определен интеграл в левой части.

Доказательство. Не ограничивая общности будем считать, что функция $f \ge 0$. Если f— простая, теорема следует из определения перенесенной меры $f_*\mu$. В случае произвольной измеримой f теорема следует из того, что f можно приблизить неубывающей последовательностью простых функций, как это делалось выше. $\square$

Пример B.25. Пусть $\Omega = \{\omega_1, \omega_2, \dots\}$ — не более, чем счетное множество, $\mathcal{F} = 2^\Omega$ и $\mu_i := \mu(\{\omega_i\})$. Тогда любая функция $f : \Omega \mapsto \mathbb{R}$ является простой, и, если она интегрируема, то

$$\int_\Omega f(\omega) \, d\mu(\omega) = \sum_j \mu_j f(\omega_j).$$

Пусть $\{a_1, a_2, \dots\}$ —множество значений функции f, а

$$\nu_i := \mu\{\omega :\ f(\omega) = a_i\} = f_*\mu(\{a_i\}).$$

Тогда, согласно теореме B.24, (в данном случае это элементарное вычисление, проведите его!)

$$\int_\Omega f(\omega)\, d\mu(\omega) = \sum_i a_i \nu_i.$$

C Список учебников

По теории вероятностей имеется масса книг. Вот некоторые из них. Основная часть списка с комментариями ниже взята из конспекта лекций В.И. Богачева 2017 года.

Я бы рекомендовал (но это мои личные предпочтения, они у всех разные!) брать книгу Кораллова-Синая, и вместе с ней листать книгу Ширяева, иначе может быть тяжеловато от суховатости изложения. Можно обойтись и одним Ширяевым, но там уж очень много всего. Если хочется еще больше примеров, можно смотреть Феллера и Тутубалина. Все или почти все книги легко скачиваются.

1. Кораллов Л.Б., Синай Я.Г. Теория вероятностей и случайные процессы. МЦНМО, М., 2013 — первая половина книги содержит компактный, четкий и интенсивный курс теории вероятностей (более объемный, чем семестровый). Наши лекции в некоторой степени следуют ей.

2. Ламперти Дж. Вероятность. Наука, М., 1973 — хорошо написанное краткое пособие для математиков.

3. Прохоров Ю.В., Пономаренко Л.С. Лекции по теории вероятностей и математической статистике. Изд-во МГУ, М., 2012 — первая половина может служить конспектом семестрового курса теории вероятностей,

4. Ширяев А.Н. Вероятность. Т. 1, 2. 3-е изд. МЦМНО, М., 2004 — весьма обстоятельное и известное современное пособие, содержащее обширный материал по многим вопросам, в том числе не входящим в обязательные программы. Книга всем хороша, кроме того, что там очень много всего, и от этого может создаться некоторое ощущение хаотичности.

5. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1, 2. Мир, М., 1984 — перевод классического трактата одного из создателей современной теории вероятностей, написанного более полувека назад, но не устаревшего и содержащего решения множества конкретных задач, включая счетно-комбинаторные. В книге рассматривается масса примеров.

6. Гнеденко Б.В. Курс теории вероятностей. 4-е изд. Наука, М., 1965 — классический учебник, написанный в период интенсивного создания современной теории более полувека назад активным участником и хорошо передающий дух теории вероятностей.

7. Тутубалин В.Н. Теория вероятностей. Академия, М., 2008 — интересное и полезное дополнительное чтение к базовому курсу, в том числе дающее впечатление о прикладной стороне дела и соотношении ее с теоретической.

   Несколько известных задачников:

8. Ширяев А.Н. Задачи по теории вероятностей. МЦМНО, М., 2006,

9. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. 2-е изд. Наука, М., 1989,

10. Кельберт М.Я., Сухов Ю.М. Вероятность и статистика в примерах и задачах. Т.1. МЦНМО, М., 2007,

11. Мешалкин Л.Д. Сборник задач по теории вероятностей. Изд-во МГУ, М., 1963,

12. Прохоров А.В., Ушаков В.Г., Ушаков Н.Г. Задачи по теории вероятностей. Наука, М., 1986.

Список литературы

[KS] Л.Б. Коралов, Я.Г. Синай, *Теория вероятностей и случайные процессы*, МЦНМО 2013.